The use of oscillatory signals in the study of genetic networks

Ovidiu Lipan *

Wing H. Wong †

Center for Biotechnology and Genomic Medicine
Medical College of Georgia
1120 15th St.,CA-4124
Augusta, Georgia 30912, USA

Departments of Statistics and Biostatistics
Harvard University
1 Oxford Street
Cambridge,Massachusetts 02138, USA

Abstract

The structure of a genetic network is uncovered by studying its response to external stimuli (input signals). We present a theory of propagation of an input signal through a linear stochastic genetic network. It is found that there are important advantages in using oscillatory signals over step or impulse signals, and that the system may enter into a pure fluctuation resonance for a specific input frequency.

The nature of a physical system is revealed through its response to external stimulation. The stimulus is imposed upon the system and its effects are then measured, Fig.1(a). This approach is widely used in biology: a cell culture perturbed with a growth factor, a heat shock etc. The data measured contain the initial information encoded into the stimulus plus the information about the intrinsic characteristics of the system. The more parameters the experimentalist can adjust to craft the perturbation stimulus, the more information about the system can be revealed. In recent years we witnessed a tremendous increase in measurement capabilities (e.g., microarray and proteomic technologies, better reporter genes). However, the success of the systems approach to molecular biology depends not only on the measurement instruments, but also on an effective design and implementation of the input stimulus, which has not been thoroughly explored. Traditionally, two types of time dependent stimuli are at work in molecular biological experiments [1], [2]. For example a step stimulus is obtained when at one instant of time a growth factor is added to the

---

*email: olipan@mail.mcg.edu

†email: wwong@stat.harvard.edu

medium, graph (a) in Fig.1(a). The stimulus from Fig.1(a) graph (b), is a superposition of two step stimuli. The investigator can control the height of the step stimulus (the concentration of the growth factor) or the time extension of the heat shock. The cells respond to these stimuli only transiently. The response is dampened after some time and becomes harder to detect it from noise. To overcome the noise, the concentration of the stimulus is typically increased to the point where the strength of the stimulus raises far above its physiological range.

We propose to implement a molecular switch at the level of gene promoter and use it to impose an oscillatory stimulus. In the absence of experimental noise, any stimulus can be used to determine the genetic network input-output properties. However, in the presence of experimental noise, the oscillatory input has many advantages: (1) the measurements can be extended to encompass many periods so the signal-to-noise ratio can be dramatically improved; (2) the measurement can start after transient effects subside, so that the data becomes easier to be incorporated into a coherent physical model; (3) an oscillatory stimulus has more parameters (period, intensity, slopes of the increasing and decreasing regimes of the stimulus) than a step stimulus. As a consequence, the measured response will contain much more quantitative information. Experimental results from neuroscience prove that oscillatory stimulus can modulate the mRNA expression level of genes. For example, c-fos transcription level in cultured neurons is enhanced 400% by an electrical stimulus at 2.5 Hz and reduced by 50% at 0.01 Hz, [3]. Also, the mRNA levels of cell recognition molecule L1 in cultured mouse dorsal root ganglion neurons change if the frequency of the electric pulses is varied. The expression level of L1 decreases significantly after 5 days of 0.1 Hz stimulation but not after 5 days of 1 Hz stimulation [4]. To extend the oscillatory approach to other type of cells, a two-hybrid assay, [5], can be used to implement a molecular periodic signal generator, Fig.1(c). The light-switch is based on a molecule (phytochrome in [5]) that is synthesized in darkness in the Q1 form. When Q1 form absorbs a red light photon (wavelength 664 nm) it is transformed into the form Q2. When Q2 absorbs a far red light (wavelength 748 nm) the molecule Q goes back to its original form, Q1. These transitions take milliseconds. The protein P interacts only with the Q2 form, recruiting thus the activation domain (AD) to the target promoter. In this position, the promoter is open and the gene is transcribed. After the desired time elapsed, the gene can be turned off by a photon from a far red light source. Using a sequence of red and far red light pulses the molecular switch can be periodically opened and closed.
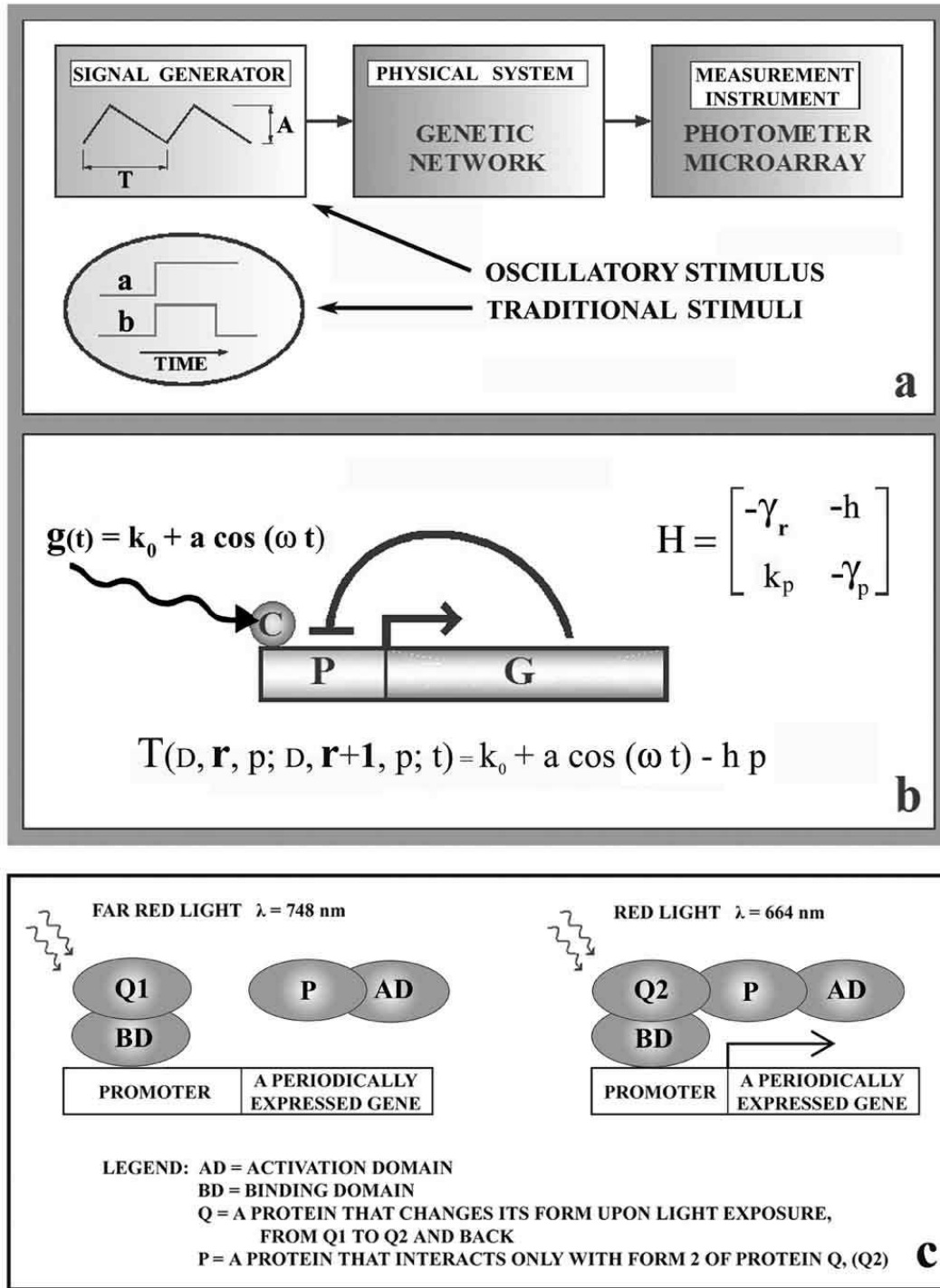
Figure 1: (a): The genetic network response depends on the type of the applied stimulus. (b): An autoregulatory network. The gene G is under the influence of a cofactor C that rhythmically modulates the activity of the promoter P. The matrix H contains the parameters that dictates the transition probabilities of the stochastic model. The transition probability per unit time from r to $r + 1$ mRNA molecules, $T(D, r, p; D, r + 1, p; t)$, is modulated by the oscillatory signal generator. The DNA, D, and the protein, p, do not change in this transition. (c), adapted from [5]. The gene is turned on with a red light pulse of wavelength $\lambda = 664$ nm. With a far red light of wavelength

3

There are four input parameters that can be varied: the period (T), the time separation between the pulses (s), and the amplitude (A) of the red and the far red pulses. The mRNA concentration profile will depend on these parameters and can be measured with a high throughput technology [6]. Protein levels will also depend on the input signal. The proteins can be recorded with 2D PAGE analysis or mass spectrometry. If one single gene product is targeted, than a real-time luminescence recording can be employed [7]. A periodic generator can be used to investigate biological system for which the mRNA and protein concentrations naturally oscillate in time. An example of such a system is the circadian clock that drives a 24-hour rhythm in living organisms from human to cyanobacteria. The core oscillator is a molecular machinery based on an autoregulatory feedback loop involving a set of key genes (Bmal, Per1,2,3, Cry1,2 , etc.) [8]. Experimental procedures used to elucidate the clock mechanism are based on measuring the circadian wheel-running behavior of mice under normal light/dark (LD) cycles or in constant darkness (dark/dark or DD) conditions. Experimental evidence demonstrates that laws of quantitative nature govern the molecular clock. For example, [9], the internal clock of cry 1 mutants have a free-running (i.e. DD conditions) period of 22.51 ± 0.06 h which is significantly lower than the period of a wild-type mice which is 23.77 ± 0.07 h. Quite opposite, a cry 2 mutant have a significantly higher period of 24.63 ± 0.06 h. In LD conditions, both mutants follow the 24 h period of the entrained light cycles. A double cry1,2 mutant is arrhythmic in DD conditions and follow a 24 h rhythm in LD conditions. To explain these experimental values we suggest using a light switchable generator to drive the expression level of cry1, 2 and measure the dynamics of transcription and the translation for the rest of the key clock genes. Another application of the periodic generator is to modulate a constitutively expressed gene by superimposing an oscillatory profile on top of its flat level. Then, the genes that show a modulation with a frequency equal to the generator's frequency will be detected by a microarray experiment. Why is this approach different from the one where a step stimulus is used? Because the frequency of the generator is *not* an internal parameter of the biological system. The genes that interact with the driven gene will be modulated by the input frequency. The rest of the genes will have different expression profiles, dictated by the internal parameters of the biological system. This point of view is supported by our findings, [6], that the circadian clock (which is an endogenous periodic signal generator) propagates its output to only $8 - 10\%$ of the transcriptome in mice peripheral tissues (liver or heart). In contrast to the oscillatory input, when

4

a step stimulus is applied, all the expression profiles are dictated by the internal parameters of the biological system. Except for the height of the step stimulus (the dose of the factor applied) there is no external parameter implemented into the input signal. As such, it is difficult to separate those genes that directly respond to the input signal and to consequently avoid artifacts. With the applications described in mind, we study the propagation of an input signal through a stochastic genetic network.

1 The response of a stochastic genetic network to an input stimulus

The effects of an oscillatory input were previously studied on specific examples using models based on differential equations [10],[11],[12]. The stochastic character is embedded into these equations as an exterior additive term. In contrast, we compute the generator's effects on the mean and fluctuation of the gene products using a stochastic model [1],[14],[2]. In this way, the generated stimulus and the noisy nature of the cell are entangled in the stochastic genetic model. For a network of n genes the state of a cell is described by the mRNA and protein molecule numbers: $q = (r_1, ..., r_n, p_1, ..., p_n)$. We assume that, during any small time interval Δt, the probability for the production of a molecule of the ith type is $(\sum_{j=1}^{2n} A_{ij} q_j + G_i(t)) \Delta t$, i.e. q_i is increased by 1 with the above probability. The function $G_i(t)$ represents the time varying input signal and modulates the mRNA production only: $G = (g_1(t), \ldots, g_n(t), 0, \ldots, 0)^T$ (the superscript T is the transposition operation that transforms G into a column vector for notational convenience in what follows). The parameter A_{ij} represents the influence of the jth type of molecule on the production rate of a molecule of the ith type. Similarly, there is a matrix of parameters Γ_{ij} governing the degradation rates of the molecules. For simplicity, we assume that the input stimulus directly affects only the production rates. The mean $\mu = \langle q \rangle$ and the covariance matrix $\nu = \langle\langle q \rangle\rangle = \langle (q - \langle q \rangle)(q - \langle q \rangle)^T \rangle$ of the state q are driven by the generator G.

The transfer of the signal from the generator through the genetic network to the output measured data is encapsulated in a set of transfer matrices. Specifically, let $H = A - \Gamma$ and denote the Laplace transforms of μ and G by $\mathcal{L}\mu$ and $\mathcal{L}G$. Here and in what follows, μ and G are represented as column vectors. The connection between the mean and the generators is given by formula
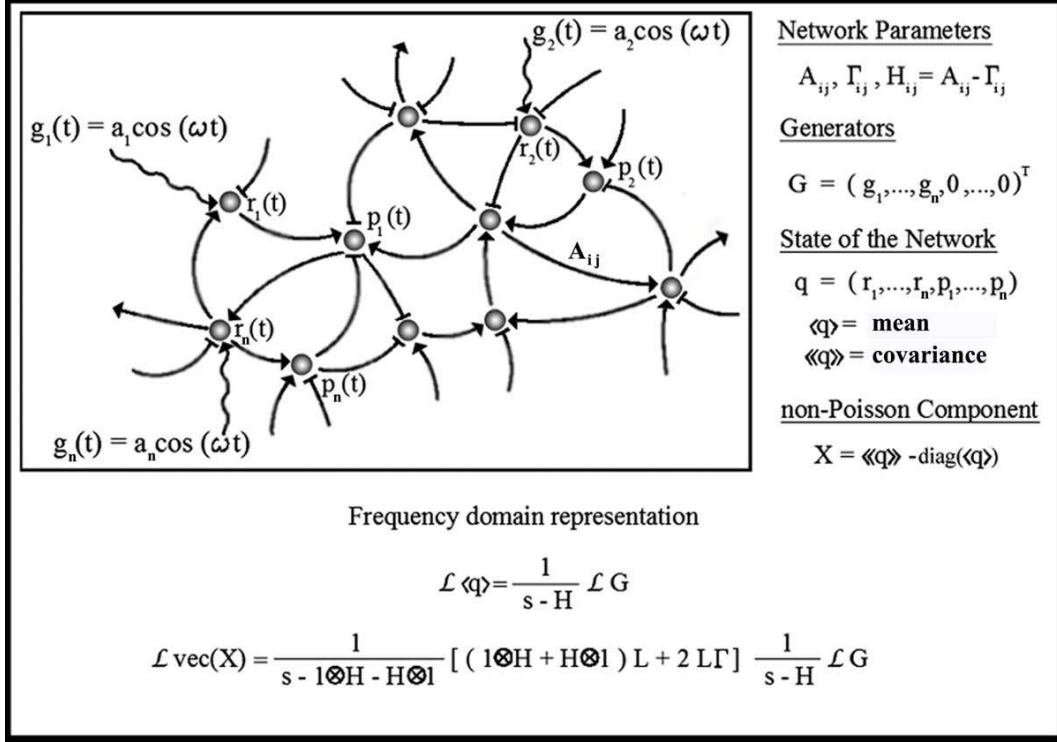
5

Figure 2: Response of a stochastic genetic network to an oscillatory input. The Laplace transform $\mathcal{L}$ change the dynamic variable from time to frequency. In the $vec(X)$ all the elements of the matrix X are arranged in a column vector.

(1) which is typical for a deterministic linear system. However, the genetic system is stochastic and the measure of the intrinsic noise is quantified by the covariance matrix ν. The effect of the stimulus generators is most transparent if we split ν in a Poisson and a non-Poisson component: $\nu = diag(\mu) + X$. Here $diag(\mu)$ represents a matrix with the components of the vector μ on its diagonal, all the other terms being zero. For a Poisson process, $X = 0$ and thus the term $diag(\mu)$ is called the Poisson component of ν. The non-Poisson component $X = \nu - diag(\mu)$ can be expressed in terms of the generators (Appendix and Supplementary Material):

$$\mathcal{L}\mu = \frac{1}{(s-H)}\mathcal{L}G . \tag{1}$$

$$\mathcal{L}vec(X) = \frac{1}{s - 1 \otimes H - H \otimes 1}\left[(1 \otimes H + H \otimes 1)L + 2L\Gamma\right]\frac{1}{s-H}\mathcal{L}G . \tag{2}$$

The $vec(X)$ is a vector constructed from the matrix X by stripping the columns of X one by one and stack them one on top of each other in $vec(X)$. We emphasize here that the time variation of the generators G in (2) can take any form and is not bounded to be periodic or a step stimulus.

6

There are 3 matrices that transfer the information from the generators to the non-Poisson component, $\mathcal{L}vec(X) = M_3 M_2 M_1 \mathcal{L}G$. The first, $M_1 = (s - H)^{-1}$, is the same as the transfer matrix for the mean. The second, $M_2 = (1 \otimes H + H \otimes 1)L + 2L\Gamma$, breaks the symmetry between the degradation and production parameters that are otherwise hidden in the matrix $H = A - \Gamma$. The $\otimes$ is the Kronecker product of two matrices. The matrix L (with elements 0 and 1) is the lifting matrix from dimension of the mean ($2n$ values) to the dimension of $vec(X)$ ($4n^2$ values). The third matrix is $M_3 = (s - 1 \otimes H - H \otimes 1)^{-1}$. If λ_i are the eigenvalues of H then all combinations $\lambda_i + \lambda_j$ are the eigenvalues of $1 \otimes H + H \otimes 1$. Thus M_3 represents the analog of M_1 in the space of covariance variables.

For a step stimulus, these eigenvalues are of primal importance: the measured signal is a superposition of components with different eigenvalues and has a complicated mathematical expression. However, for a periodic stimulus, the frequency of the external generator is the important parameter. This frequency is fixed by the experimentalist not by the biological system. Only the phase and the amplitude of the output signal depends on the system's eigenvalues and the mathematical form is less cumbersome then for the step stimulus. The input-output relations, (1) and (2), were derived from the Master Equation written for the probability of the states of the genetic network. Thus we must specify the initial conditions for the probability of the states. These conditions refer here to states for which one molecular component vanishes ($q_i = 0$, for one i). The input-output relations, (1) and (2), are independent of these boundary states if the Γ matrix is diagonal. A diagonal Γ matrix was used in [1] and we will use it also in the example that follows. Tools developed in the field of System Identification can be used to create models for the networks under study, [16]. The difference between the System Identification classical models and a genetic network is that the later is a stochastic process by nature, whereas the former are deterministic models with a superimposed noise from external sources. However the formulas that describe the relations between the mean and covariance of the stochastic process and the input signals, (1) and (2), are of the same general nature as those used in System Identification Theory, [16]. In the next section we will use (1) and (2) to analyze one of the most fundamental regulatory motifs in a genetic network: an autoregulatory gene that acts upon itself through a negative feedback, [4],[18],[19]. The fluctuation can drive this biological system out of its equilibrium state,[20].

2 Fluctuation resonance

Four parameters characterize the system: the feedback strength $A_{12} = -h$; the translation rate $A_{21} = k_p$, and two degradation rates, $\Gamma_{11} = \gamma_r$, $\Gamma_{22} = \gamma_p$. The gene regulation is under the control of its own protein product and the protein activity is modulated by a cofactor. The cofactor is driven by a periodic light switchable generator $g(t) = k_0 + a\cos(\omega t)$, Fig.1(a). Before the generator is applied, the transcription rate is equal with k_0 and the system is in a steady state. Through the transfer matrices, (1) and (2), the light generator will impose a periodic evolution of the mean and covariance matrix for mRNA and protein product. We denote the mean mRNA by $\langle r(t) \rangle$ and the mean number of protein by $\langle p(t) \rangle$. We will concentrate on the protein number in what follows. After the transients are gone, $\langle p(t) \rangle = P_0 + P_1 e^{i\omega t} + P_1^* e^{-i\omega t}$, that is the protein number will oscillate with an amplitude P_1 on top of a baseline P_0; here $*$ represents complex conjugation. The fluctuation of the protein number, $\langle\langle p(t) \rangle\rangle$, differs from the mean number by a quantity that we denoted by $X_{pp}(t)$: $\langle\langle p(t) \rangle\rangle = \langle p(t) \rangle + X_{pp}(t)$. For a pure Poisson process, $\langle\langle p(t) \rangle\rangle = \langle p(t) \rangle$. Thus the term $X_{pp}(t)$ represents the deviation from a Poisson process. If there is some information about the genetic system that can be uncovered by measuring not only the mean but also the covariance matrix, then this information is hidden only in the non-Poisson component $X_{pp}(t)$. The quantity $X_{pp}(t)$ is not interesting only from a statistical point of view but also from a dynamical one. The equation for the time evolution of $\langle\langle p(t) \rangle\rangle$ takes its most simple form if it is written for $X_{pp}(t)$. That is, the time dependence of the mean value must be subtracted from the time evolution of $\langle\langle p(t) \rangle\rangle$. Similar to the mean value, the non-Poisson component of the fluctuation will oscillate in time, $X_{pp}(t) = X_{p,0} + X_{p,1} e^{i\omega t} + X_{p,1}^* e^{-i\omega t}$ with complex amplitude $X_{p,1}$. The relative strength of the fluctuation versus the mean value can be described using the Fano factor, [1]: $\langle\langle p(t) \rangle\rangle / \langle p(t) \rangle = 1 + X_{pp}(t)/\langle p(t) \rangle$. For oscillatory inputs, the response of the network is best described in frequency domain rather than in time. In frequency domain, as an analog of the Fano factor we consider the ratio of the amplitude of $X_{pp}(t)$ versus the amplitude of $\langle p(t) \rangle$.

$$\frac{|X_{p,1}|}{|P_1|} = \left(4\,k_p{}^2 \frac{\left(\omega^2 + (h - \gamma_p)^2\right)\left(\omega^2 + 4\,\gamma_r{}^2\right)}{\left((\omega^2 - 4\,\omega_0{}^2)^2 + 4\,\omega^2\omega_1{}^2\right)(\omega^2 + \omega_1{}^2)} \right)^{1/2} . \tag{3}$$

Here $\omega_1 = \gamma_r + \gamma_p$. The complex amplitudes $X_{p,1}$ and P_1 depend on the input frequency and therefore resonance phenomena can be detected in the system. If the light switchable generator
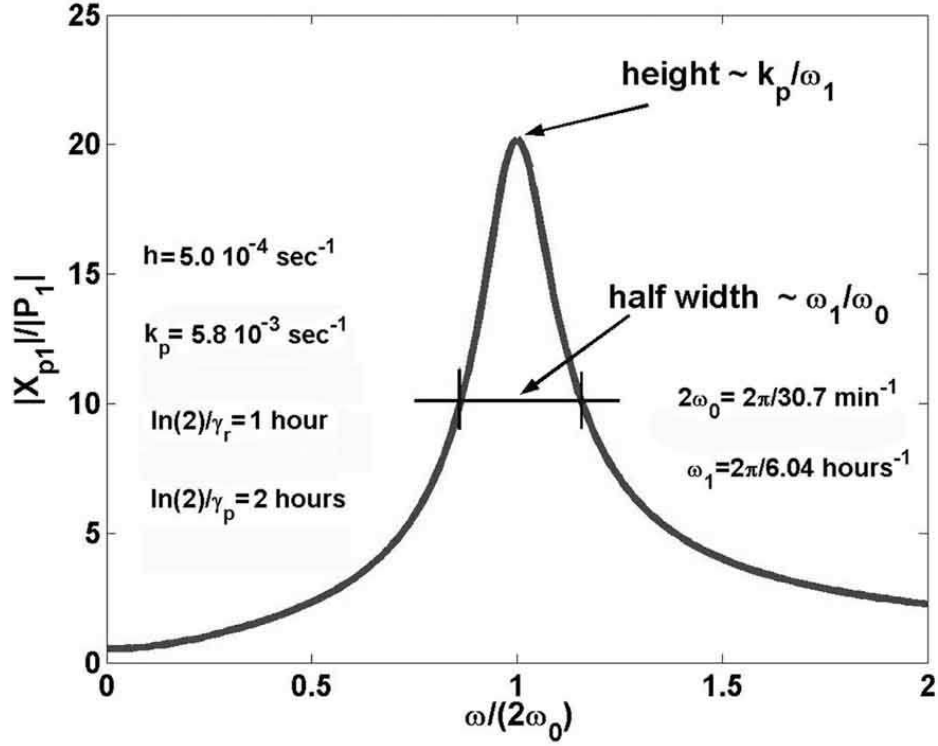
Figure 3: Fluctuation resonance. The amplitude $X_{p,1}$ of the non-Poisson component is much higher than the amplitude of the mean protein number, P_1, at $\omega = 2\omega_0$.

oscillates with double the natural frequency $\omega_0^2 = hk_p + \gamma_r\gamma_p$, that is , $\omega = 2\omega_0$ we find a state of resonance for fluctuation and not for the mean, Fig.3.

For $\omega = 2\omega_0$ the system will be in a pure fluctuation resonance. In such a situation the molecular noise can drive the cell out of its equilibrium state, which can have dramatic consequence on the cell fate. Our model being linear cannot cover the entire phenomena that accompanies a system whose state is close to resonance. However, a linear model suggest the existence of pure fluctuation resonance. At fluctuation resonance, the deviation from a Poissonian process is high. The oscillation amplitude for protein fluctuation is much greater then the amplitude of the mean. Experimental results [21] show that typical values for the ratio k_p/γ_r are 40 for $lacZ$ and 5 for $lacA$. This suggests that there are natural conditions for a strong height fluctuation resonance, Fig.3. However, for a sharp fluctuation resonance (small half width), we need $h > \gamma_r$ or γ_p, a condition that does not appear in all genetic networks. It is with the help of the experimental study that we will clarify why some biological systems can sustain fluctuation resonance and others not. Beside resonance,

the frequency response provides other insights into the structure of the autoregulatory system. The parameters of the system can be read out from the measured data. The frequency response of the mean values behave like the response of a classical linear system to input signals. The new aspects are those related to fluctuations. Like $X_{pp}(t)$ and $X_{rr}(t)$, the correlation coefficient between the mRNA and protein number will oscillate in time: $X_{rp}(t) = X_{rp,0} + X_{rp,1}e^{i\omega t} + X^*_{rp,1}e^{-i\omega t}$ with amplitude $X_{rp,1}$. Taking the ratios of the amplitudes: $|X_{rp,1}|^2/|X_{r,1}|^2 = (1/4h^2)\omega^2 + \gamma_r^2/h^2$, $|X_{rp,1}|^2/|X_{p,1}|^2 = (1/4k_p^2)\omega^2 + \gamma_p^2/k_p^2$, we observe that all four parameters of the system can be estimated from the slopes and the intercepts of the above ratios as a function of ω^2. Detail formulas for each amplitude are given in the Supplementary Material.

3   The spectrum, the experimental noise and the importance of the input stimulus

We described the use of a periodic signal to decipher a genetic network. Traditionally, a step stimulus is employed in biology for pathway detection (i.e., adding a growth factor to the culture). From the response to a step stimulus we can extract, in principle, the parameters of the system. The natural question is then: why should we generate a periodic stimulus when there is already a step stimulus in use? Seeking an answer, we notice that the measured data in our studied example can be expressed as a sum of exponentially decaying functions, $e^{-\lambda t}$, if a step stimulus was used (Supplementary Material). For a periodic input, the response contains only exponentials with imaginary argument, $e^{i\omega t}$. Mathematically, the main difference between exponentials with real arguments, $e^{-\lambda t}$, and those with imaginary arguments, $e^{i\omega t}$, is that with the former we can not form an orthogonal basis of functions whereas such a basis can be formed with the later. If we depart from our example, we can say that in general, the response of the network to a step input will be a sum of components which are not orthogonal on each other. The time dependance of these nonorthogonal components can be more complex than an exponential function; they can contain polynomials in time or decaying oscillations, depending on the position in the complex plane of eigenvalues of the transfer matrix H. Contrary, the permanent response obtained from a periodic input is a sum of Fourier components that form an orthogonal set. Orthogonal components are much more easy to separate than nonorthogonal ones. This mathematical difference explains the advantage of using

oscillatory inputs. However, an argument can be made that increasing the number of replicates will be enough to recover the step response form noise. In what follows we study how many replicates we need to successfully fight the experimental noise. We will show that we need fewer replicates if the genetic network is probed with an oscillatory generator than with a step signal. To keep the argument simple, we will study the difficulty of separating nonorthogonal components for a network for which the response to a step stimulus is a sum of decaying exponentials. The argument can be extended to other types of nonorthogonal components, but this line of thought falls out of the scope of this paper. The measured data being a superposition of exponential terms can be written as:

$$f(t) = \int_{x_1}^{x_2} S(x) K(xt) \, dx \, , \tag{4}$$

with $K(xt) = e^{-ixt}$ for the periodic response and $K(xt) = e^{-xt}$ for the step stimulus. The spectral function $S(x)$ depends on the network's parameters and on the type of the input signal. For example, the spectrum of the autoregulatory system for a periodic input is $S(x) = S_0\delta(x) + S_1\delta(x - i\omega) + S_1^*\delta(x + i\omega)$, where $\delta(x)$ is the Dirac delta function. The coefficients S_0, S_1 take specific values if the spectrum refers to mean mRNA, proteins or their correlations. For example, for the protein fluctuation:

$$S_0 = X_{p,0} = \frac{k_p{}^2 k_0 \left(\gamma_p - h\right) \gamma_r}{\omega_0{}^4 \omega_1} \, , \tag{5}$$

$$S_1 = X_{p,1} = \frac{ia\left(-i\gamma_p + \omega + ih\right)\left(\omega - 2\,i\gamma_r\right) k_p{}^2}{\left(\omega^2 - \omega_0{}^2 - i\omega\,\omega_1\right)\left(\omega^2 - 2\,i\omega\,\omega_1 - 4\,\omega_0{}^2\right)\left(\omega - i\omega_1\right)} \, . \tag{6}$$

Detailed description of the spectrum for an autoregulatory network is given in the Section 5 of the Supplementary Material. For oscillatory inputs that are not pure cosine function and for more complicated networks, the spectrum is more complex, but still is connected with the measured data like in (4). The spectrum $S(x)$ carries information about the parameters of the genetic network and it can be recovered from the data $f(t)$. The network's parameter can be estimated from the spectrum once a model of the network is chosen. Our goal is to show that the spectrum obtained from an oscillatory input signal is much less distorted by the experimental noise than the spectrum obtained from a step input. Laboratory measurements are samples of $f(t)$ at N discrete time points. Given a finite number N of measured data points, $f_1, \cdots, f_N$, the spectrum for the periodic case $S(x)$ can only be approximated as a weighted sum of N terms, (Supplementary Material): $S(x) = \sum_{k=1}^{N}(s_k + \epsilon_k/\beta_k)\Theta_k(x)$. Each term, $(s_k + \epsilon_k/\beta_k)\Theta_k(x)$, contain a function $\Theta_k(x)$ that do

11

not depend on the measured data, and the weights $s_k + \epsilon_k/\beta_k$ that are computed from the measured data $f_1, \cdots, f_N$. In the absence of experimental noise, $\epsilon_k = 0$, all N coefficients s_k can be computed from the measured data. When experimental noise is present, $\epsilon_k \neq 0$, what we compute from measured data is $s_k + \epsilon_k/\beta_k$, and we cannot separate s_k from it because we do not know the actual value for ϵ_k. The best we can do is to use only those terms for which $s_k > \epsilon_k/\beta_k$, so the effect of the distortion on s_k is not large. Unfortunately, the distortion increases as β_k goes smaller, which actually happens when k increases. A term can be recovered from noise if $\beta_k^{-1} < s_k/\epsilon_k$. Usually, this relation is valid for $k = 1 \cdots J_p$, with J_p being the last term that can be recovered. A similar relation holds for the exponential case, with α_k instead of β_k and J_e instead of J_p. It is desirable that both cutoffs (J_p, J_e) be as close as possible to the number of sampled points, N. The striking difference between the two cases is that the cutoff J_p is much larger then the cutoff J_e. This is a consequence of the fact that the numbers α_k decrease exponentially to 0, [5], whereas β_k stays close to 1 for many k before eventually dropping close to zero, [6]. This huge difference between α_k and β_k has its origin in the fact that the set of functions of time, $exp(-\lambda t)$, indexed by λ, do not form an orthogonal set, whereas the functions $exp(i\omega t)$, indexed by ω, are orthogonal. In theory, however,
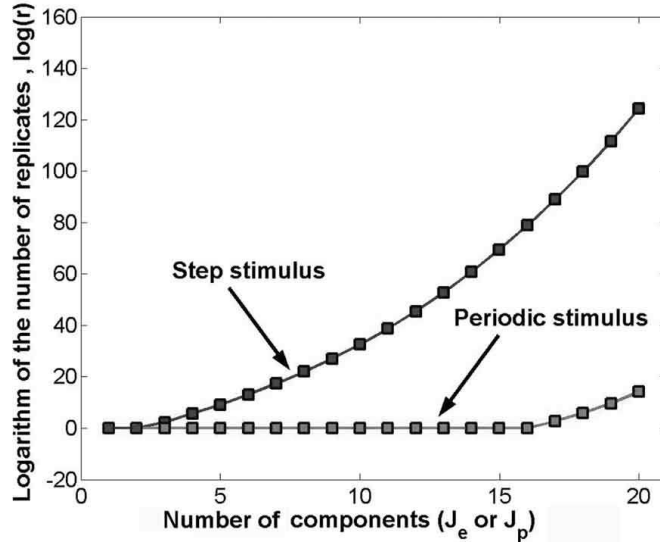

Figure 4: How many replicates we need to recover a given spectral component.

we can still hope that a step stimulus can deliver good estimates if the noise ϵ_k is reduced using r replicates $(\epsilon_k \rightarrow \epsilon_k/\sqrt{r})$. This is not the case. Fig.4 represents the number of replicates needed to recover the component J_e or J_p if the Signal to Noise Ratio is 10 ($SNR \equiv s_{J_e}/\epsilon_{J_e} = s_{J_p}/\epsilon_{J_p} = 10$).

The number of replicates grows very fast in the exponential case (for $SNR = 10$ and $N = 20$ we need 269 replicates for the 4th spectral component), whereas in the periodic case, the number of replicates stays low for many spectral components (only for the 17th component it raises to 14, with $SNR = 10$ and $N = 20$).

4   Conclusions and Discussions

We studied the response of a linear stochastic genetic network to an input stimulus (signal). We provide a general formula that relates the mean and covariance matrix of mRNAs and proteins to the input generators. The particular type of periodic signals was studied in detail for an autoregulatory system. We found that fluctuation resonance can manifest in such systems. Besides interesting physical phenomena that can be detected using a periodic signal, the oscillatory input is useful for experimental noise rejection. We compared two experimental designs: one that uses a step stimulus as a perturbation and another one that uses a periodic input. We concluded that the response of the genetic network to a periodic stimulus is much easier to be detected from noise than the response of the same network to a step stimulus. This conclusion applies whenever the response of the network to an oscillatory input is a sum of Fourier components. This can be the case for many nonlinear networks. However, the input-output relations, (1) and (2), applies only to a linear stochastic model. A linear model is a good approximation around a steady state of the genetic network. A genetic network is a nonlinear system and can have several steady states. If the signal generator does not vary in time, the genetic network will be characterized by one of these steady states. When the signal generator starts to oscillate with an amplitude that doesn't drive the network far away from its steady state, the linear model is a good approximation. For large amplitudes, the nonlinear effects start to be important, and at some values of the generator's amplitude, the network will jump close to a different steady state. Such nonlinear behaviors can not be described by a linear model. Also, the parameters that describe the network are supposed to be constant in time. This approximation is valid if the changes in the network parameters are slow with respect to the changes produced by the oscillatory input signals. The input frequency should be chosen so that the system can be considered with constant coefficients for the elapsed time of measurements. Also, the period of oscillations must be less than the trend effects due to growth,

apoptosis etc. Beside biological effects that span large intervals of time, experimental artifacts, like medium evaporation, can superimpose a trend on the measured profile. The input period should be less than the time characteristics of these trends. This will impose a limit for the lower range of the input frequencies. The response to oscillations depends also on the time characteristics of the system under study. If the system has a high damping factor, the high frequencies will be strongly attenuated and the output signal is not measurable. With all these restrictions, the experimentalist still has the freedom to work in a frequency band, a freedom not present in the step stimulus.

A different line of thought emerges when it comes to analyzing if the oscillatory method can be scaled to large networks. Experimentally, using high throughput measurements (microarray and proteomic tools) a large set of gene products can simultaneously be measured. The experimentalist is searching for a pathway that is controlled by a gene. Using oscillatory signals to stimulate the desired gene, the time variation of the downstream genes will contain in its spectrum the input frequency and so these genes will be detected. Moving the signal generator along the pathway, more and more local patches of the network will be uncovered. The global view on the network will consist of all these patches connected together. The theoretical framework for connecting a set of patches is unclear to us at present. Experimentally however, we verified that a source of oscillations propagates into a large genetic network,[6]. Specifically, a microarray experiment was conducted on mice entrained for two weeks on a 24 hours period of light-dark signals. The periodic input signal was not implemented at the level of gene promoter; it was an exterior periodic source of light that entrained the internal clock of the cell. After entrainment, and in complete darkness, the output signals (mRNA) were measured every 4 hours for 2 days using and Affymetrix platform. From about 6000 expressed genes in heart, about 500 showed a mRNA that oscillates with a 24 hours period. Same results were reported in [24]. The next step is to implement the generator at the gene promoter level, and measure the spread of the input signal into the network.

Given the advantages of a periodic stimulus presented above, we believe that the experimental implementation of a periodic generator at promoter level will prove fruitful in the study of genetic networks.

Appendix

The genetic network is described by a linear stochastic network $[1], [14], [2]$. The network is driven using signal generators placed inside the promoters of a subset of genes that are part

of the network. For a gene we will denote by (D, r, p) the number of DNA, mRNA and protein molecules respectively per cell. We consider r, p variables but D constant and we normalize it to $D = 1$. The state of a cell that contains n active genes is specified by: $\tilde{q} = (D_1, D_2, ..., D_n, r_1, r_2, ..., r_n, p_1, p_2, ..., p_n)$. The genetic state is changing in time; for a short transition time, dt, only one $\tilde{q}_i$ changes its value and this new value can be either $\tilde{q}_i + 1$ or $\tilde{q}_i - 1$. We consider in this paper a linear stochastic genetic network characterized by the following transition probabilities: $T(\tilde{q}; \tilde{q} + 1_i; t) = \sum_{j=1}^{M} \tilde{A}_{ij} \tilde{q}_j dt$, $T(\tilde{q}; \tilde{q} - 1_i; t) = \sum_{j=1}^{M} \tilde{\Gamma}_{ij} \tilde{q}_j dt$. Here $\tilde{q}$ is the initial state and 1_i is a vector of length M with all elements 0 except the one in the position i which is 1. The time variation of the generators that drive the genes' expressions are encapsulated in the matrix $\tilde{A}_{ij}$ which governs the production of different molecules. The matrices $\tilde{A}_{ij}$ and $\tilde{\Gamma}_{ij}$ consist of four submatrices, corresponding to splitting the state $\tilde{q}$ in two subgroups. One subgroup contains only the DNA states $(D_1, \cdots, D_n)$ and the other subgroup contains the protein and mRNA states $q = (r_1, r_2, ..., r_n, p_1, p_2, ..., p_n)$,

$$\tilde{A} = \begin{bmatrix} 0 & 0 \\ Gen & A \end{bmatrix}, \quad \tilde{\Gamma} = \begin{bmatrix} 0 & 0 \\ 0 & \Gamma \end{bmatrix}. \tag{7}$$

The generator submatrix Gen has a special form. It is a $2n \times n$ matrix and locates the position of the generators in the genetic network: $Gen_{ij} = g_i(t)\delta_{ij}$, $i = 1 \ldots 2n$, $j = 1 \ldots n$. Each gene promoter is driven by one generator $g_i(t), i = 1, \ldots, n$, which will influence the mRNA production of gene i. The same mRNA production can be influenced by the protein concentration, and this feedback effect is described by the elements of the $2n \times 2n$ matrix A, (7). The structure of the matrix A is a consequence of the topology of the genetic network. The equation for the probability $P(\tilde{q}, t)$ of the network to be in the state $\tilde{q}$ at time t is: $\partial P(\tilde{q}, t)/\partial t = \sum_{i=1}^{M} (E_i^- - 1) \sum_{k=1}^{M} \tilde{A}_{ik} \tilde{q}_k P(\tilde{q}, t) + \sum_{i=1}^{M} (E_i^+ - 1) \sum_{k=1}^{M} \tilde{\Gamma}_{ik} \tilde{q}_k P(\tilde{q}, t)$, where the shift operators $E_i^{\pm}$ are given by $E_i^{\pm} P(\tilde{q}, t) = P(\tilde{q}_1, ..., \tilde{q}_i \pm 1, ..., \tilde{q}_M)$.

We need the time evolution equations for mRNAs and proteins: $\mu_i = <q_i>$ and $\nu_{ij} = <q_i q_j> - <q_i><q_j>$, $i, j = 1, \ldots 2n$. In matrix notation, for the column vector μ and for the matrix X with elements given by $X_{ij} = \nu_{ij} - \delta_{ij}\mu_i$ we obtain:

$$\frac{d}{dt}\mu = H\mu + G, \tag{8}$$

$$\frac{d}{dt}X = HX + XH^T + H\,diag(\mu) + diag(\mu)H^T + 2diag(\Gamma\mu). \tag{9}$$

Here H^T is the transpose matrix of $H = A - \Gamma$ and $diag(\mu)$ has nonzero elements only on the principal diagonal: $diag(\mu)_{ij} = \delta_{ij}\mu_i$. Using the Laplace transform, the solution to (8) is (1). The second equation, (9), is a matrix equation. To solve this equation we first transform it to an equation were the unknown is a column vector. The transformation needed is $X \mapsto vec(X)$, where the column vector $vec(X)$ contains the columns of the matrix X one on top of the next one, starting with the first column and ending with the last column. The vec mapping has the useful property that $vec(HX) = (1 \otimes H)vec(X)$, $vec(XH) = (H^T \otimes 1)vec(X)$, were 1 is the unit matrix and $A \otimes B$ is the tensor product of two matrices A and B. The column vector $vec(diag\,(\mu))$ can be expressed in terms of the column vector μ: $vec(diag\,(\mu)) = L\mu$, were L is a lift matrix from a space of dimension of μ to the square of this dimension: $L = (P_1, \ldots, P_{2n})^T$, $(P_k)_{ij} = \delta_{ik}\delta_{jk}$. The solution to (9) takes the form (2).

Acknowledgements

References

[1] Gardner, T.S., di Bernardo, D., Lorenz, D. & Collins, J.J.(2003) *Science* **301**, 102-105.

[2] Vance, W., Arkin, A. & Ross, J. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 5816-5821.

[3] Sheng, H.Z., Fields, R.D. & Nelson, P.G. (1993) *J. Neurosci. Res.* **35**, 459-467.

[4] Itoh, K., Stevens, B., Schachner, M. & Fields, R.D. (1995) *Science* **270**, 1369-1372.

[5] Shimizu-Sato, S., Huq, E., Tepperman, J.M. & Quail, P.H. (2002) *Nature Biotechnology* **20**, 1041-1044.

[6] Storch, K.F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F.C., Wong, W.H. & Weitz, C.J. (2002) *Nature* **417**, 78-83.

[7] Izumo, M., Johnson, C.H. & Yamazaki, S. *Proc Natl Acad Sci USA* (2003) **100**, 16089-16094.

[8] Reppert, S. M. & Weaver, D. R. (2002) *Nature* **418**, 935-941.

[9] van der Horst, G.T.J., Muijtjens, M., Kobayashi, K., Takano, R., Kanno, S., Takao, M., de Wit, J., Verkerk, A., Eker, A.P.M., van Leenen, D. et al. (1999) *Nature* **398**, 627-630.

[10] Smolen, P., Baxter, D. A. & Byrne, J. H. (1998) *Am. J. Physiol.* **43**, C531 .

[11] Hasty, J., Dolnik, M., Rottschafer, V. & Collins, J.(2002) *Phys. Rev. Lett.* **88** (14), art. no. 148101 .

[12] Simpson, M.L., Cox, C.D. & Sayle, G.S. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 4551-4556.

[13] Thattai, M. & van Oudenaarden, A. (2001) *Proc. Natl Acad. Sci.USA* **98**, 8614-8619.

[14] Swain, P.S., Elowitz, M.B. & Siggia, E.D. (2002) *Proc. Natl Acad. Sci.USA* **99**, 12795-12800 .

[15] van Kampen, N. G. (1992) *Stochastic Processes in Physics and Chemistry* (NorthHolland, Amsterdam).

[16] Ljung, L. *System Identification - Theory for the User*, (1999), Prentice Hall, Upper Saddle River, N.J. 2nd edition

[17] Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I. et al. (2002) *Science* **298**, 799-804.

[18] Becskei, A. & Serrano, L. (2000) *Nature* **405**, 590-593.

[19] Rosenfeld, N., Elowitz, M. & Alon, U. (2002) *J. Mol. Biol.* **323**, 785-93.

[20] Isaacs, F., Hasty, J., Cantor, C. & Collins, J.(2003) *Proc Natl Acad Sci USA* **100**, 7714-7719.

[21] Kennell, D. & Riezman, H.(1977) *J. Mol. Biol.* **114**, 1-21.

[22] Bertero, M., Brianzi, P. & Pike, E.R.(1985) *Proc.R.Soc.Lond.A* **398**, 23-44.

[23] Slepian, D. (1983) *SIAM Review* **25**, 379-393.

[24] Panda, S., Antoch, M.P., Miller, B.H., Su, A.I., Schook, A.B., Straume, M., Schultz, P.G., Kay, S.A., Takahashi, J.S. & Hogenesch, J.B.(2002) *Cell* **109**, 307-320.

A Supporting Material

A.1 Derivation of the time evolution equations for the mean and fluctuation driven by signal generators

The genetic state $\tilde{q} = (D_1, D_2, ..., D_n, r_1, r_2, ..., r_n, p_1, p_2, ..., p_n)$ is changing in time; let the state be $\tilde{q}_{initial}$ at time t_1 and $\tilde{q}_{final}$ at a later time t_2. The probability of transition from the initial to the final state is, in the most general case, a function of the initial state, the final state and the times of transitions: $T(\tilde{q}_{initial}; \tilde{q}_{final}; t_1, t_2)$. Following a common hypothesis, the transition probability is proportional with the transition time $t_2 - t_1$ if this is very short ($t_2 = t_1 + dt$ with dt an infinitesimal small quantity). The transition time being short only one $\tilde{q}_i$ changes its value and this new value can be either $\tilde{q}_i + 1$ or $\tilde{q}_i - 1$. We consider in this paper a linear stochastic genetic network characterized by the following transition probabilities: $T(\tilde{q}; \tilde{q}+1_i, ; t, t+dt) = \sum_{j=1}^{M} \tilde{A}_{ij} \tilde{q}_j dt$, $T(\tilde{q}; \tilde{q}-1_i; t, t+dt) = \sum_{j=1}^{M} \tilde{\Gamma}_{ij} \tilde{q}_j dt$. Here $\tilde{q}$ is the initial state and 1_i is a vector of length M with all elements 0 except the one in the position i which is 1. The equation for the probability of the network to be in the state $\tilde{q}$ at time t, $P(\tilde{q}, t)$, is then, [1] [2],:

$$\frac{\partial}{\partial t} P(\tilde{q}, t) = \sum_{i=1}^{M} \left(E_i^- - 1 \right) \sum_{k=1}^{M} \tilde{A}_{ik} \tilde{q}_k P(\tilde{q}, t) + \sum_{i=1}^{M} \left(E_i^+ - 1 \right) \sum_{k=1}^{M} \tilde{\Gamma}_{ik} \tilde{q}_k P(\tilde{q}, t) , \tag{10}$$

where the shift operators $E_i^{\pm}$ are given by

$$E_i^{\pm} P(\tilde{q}, t) = P(\tilde{q}_1, ..., \tilde{q}_i \pm 1, ..., \tilde{q}_M). \tag{11}$$

We need to obtain the time evolution equations for $< \tilde{q}_\alpha >$ and $< \tilde{q}_\alpha \tilde{q}_\beta >$,

$$< \tilde{q}_\alpha > \equiv \sum_{\tilde{q}=0}^{\infty} \tilde{q}_\alpha P(\tilde{q}, t) , \tag{12}$$

$$< \tilde{q}_\alpha \tilde{q}_\beta > \equiv \sum_{\tilde{q}=0}^{\infty} \tilde{q}_\alpha \tilde{q}_\beta P(\tilde{q}, t) . \tag{13}$$

The easiest way to follow the computations is to use the z-transform of a function, defined by:

$$\mathbf{Z}(P(\tilde{q}, t)) = \sum_{\tilde{q}_1=0,...,\tilde{q}_M=0}^{\infty} z_1^{\tilde{q}_1} \ldots z_M^{\tilde{q}_M} P(\tilde{q}, t) . \tag{14}$$

The argument z of the z-transform will be displayed using the notation

18

$$F\left(z,t\right) = \mathbf{Z}(P\left(\tilde{q},t\right)). \tag{15}$$

The quantities of interest are related with the z-transform through

$$F_\alpha \quad = \quad <\tilde{q}_\alpha>, \tag{16}$$

$$F_{\alpha\beta} \quad = \quad <\tilde{q}_\alpha\tilde{q}_\beta> -\delta_{\alpha\beta}<\tilde{q}_\alpha> .$$

where $\delta_{\alpha\beta}$ is the Kronecker delta symbol which is 0 if $\alpha \neq \beta$ and 1 if $\alpha = \beta$ and

$$F_\alpha \quad = \quad \frac{\partial}{\partial z_\alpha}F\left(z,t\right)|_{z_i=1,\ i=1...M}, \tag{17}$$

$$F_{\alpha\beta} \quad = \quad \frac{\partial}{\partial z_\alpha\partial z_\beta}F\left(z,t\right)|_{z_i=1,\ i=1...M}. \tag{18}$$

The derivatives of the z-transform are not directly related to the covariance matrix:

$$\nu_{\alpha\beta} =<\tilde{q}_\alpha\tilde{q}_\beta> - <\tilde{q}_\alpha><\tilde{q}_\beta> . \tag{19}$$

However, the covariance matrix can be easily expressed in terms of the z-transform variables:

$$\nu_{\alpha\beta} = F_{\alpha\beta} - F_\alpha F_\beta + \delta_{\alpha\beta}F_\alpha . \tag{20}$$

The equation for F can be obtained by taking the z-transform of the master equation (10) using the following rules:

$$\mathbf{Z}\left(E_i^+ P(\tilde{q},t)\right) \quad = \quad z_i^{-1}\mathbf{Z}\left(P(\tilde{q},t)\right) - z_i^{-1}\mathbf{Z}\left(P(\tilde{q},t)|_{\tilde{q}_i=0}\right), \tag{21}$$

$$\mathbf{Z}\left(E_i^- P(\tilde{q},t)\right) \quad = \quad z_i\mathbf{Z}\left(P(\tilde{q},t)\right), \tag{22}$$

$$\mathbf{Z}\left(\tilde{q}_i P(\tilde{q},t)\right) \quad = \quad z_i\partial_{z_i}\mathbf{Z}\left(P(\tilde{q},t)\right). \tag{23}$$

$$\tag{24}$$

If the degradation matrix Γ is diagonal [1], then the probability $P(\tilde{q},t)|_{\tilde{q}_i=0}$ of the state with a missing molecular specie will not be part of the the equation for the z-transform. Indeed, the boundary term in the z-transform of $E_i^+\Gamma_{ii}\tilde{q}_i P(\tilde{q},t)$ will vanish for $\tilde{q}_i = 0$. For a non-diagonal Γ matrix, we obtain the same equation if we work with natural boundary conditions, that is $P(\tilde{q},t) = 0$ if $\tilde{q}_i = 0$ for one i from the set $1...M$. The majority of the genetic networks will not

obey the natural boundary conditions. However, the final results are the same for (i) a non-diagonal Γ matrix with natural boundary conditions and (ii) a diagonal Γ matrix with no restriction imposed on the boundary. For the sake of the symmetry of the computations, we will derive the results for a general Γ matrix and natural boundary conditions, and use a diagonal Γ matrix when we will study the behavior of a genetic network.

The equation for the z-transform now reads:

$$\frac{\partial}{\partial t}F\left(z,t\right) = \sum_{i=1}^{M}\left(z_i - 1\right)\sum_{k=1}^{M}\tilde{A}_{ik}z_k\frac{\partial}{\partial z_k}F\left(z,t\right) + \sum_{i=1}^{M}\left(z_i^{-1} - 1\right)\sum_{k=1}^{M}\tilde{\Gamma}_{ik}z_k\frac{\partial}{\partial z_k}F\left(z,t\right) \ . \tag{25}$$

Take the derivative of these equation with respect to z_α

$$\frac{\partial^2}{\partial t\partial z_\alpha}F\left(z,t\right) = \sum_{i,k=1}^{M}\tilde{A}_{ik}\left(\delta_{i\alpha}z_k\frac{\partial}{\partial z_k}F\left(z,t\right) + \left(z_i - 1\right)\delta_{k\alpha}\frac{\partial}{\partial z_k}F\left(z,t\right) + \left(z_i - 1\right)z_k\frac{\partial^2}{\partial z_k\partial z_\alpha}F\left(z,t\right)\right) +$$

$$\sum_{i,k=1}^{M}\tilde{\Gamma}_{ik}\left(-z_i^{-2}\delta_{i\alpha}z_k\frac{\partial}{\partial z_k}F\left(z,t\right) + \left(z_i^{-1} - 1\right)\delta_{k\alpha}\frac{\partial}{\partial z_k}F\left(z,t\right) + \left(z_i^{-1} - 1\right)z_k\frac{\partial^2}{\partial z_k\partial z_\alpha}F\left(z,t\right)\right) \ .$$

Introducing $z_i = 1, i = 1...M$ we obtain the equation for the time evolution of the mean values:

$$\frac{d}{dt}F_\alpha = \sum_{k=1}^{M}(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k})F_k \ . \tag{26}$$

For the second moments we continue to take derivatives of (25):

$$\frac{\partial^3}{\partial t\partial z_\alpha\partial z_\beta}F\left(z,t\right) = \sum_{i,k=1}^{M}\tilde{A}_{ik}(\delta_{i\alpha}\delta_{k\beta}\partial_k F + \delta_{i\alpha}z_k\partial_{k\beta}F + \delta_{i\beta}\delta_{k\alpha}\partial_k F + \left(z_i - 1\right)\delta_{k\alpha}\partial_{k\beta}F +$$

$$\delta_{i\beta}z_k\partial_{k\alpha}F + \left(z_i - 1\right)\delta_{k\beta}\partial_{k\alpha}F + \left(z_i - 1\right)z_k\partial_{k\alpha\beta}F) +$$

$$\sum_{i,k=1}^{M}\tilde{\Gamma}_{ik}(2\,z_i^{-3}\delta_{i\beta}\delta_{i\alpha}z_k\partial_k F - z_i^{-2}\delta_{k\beta}\delta_{i\alpha}\partial_k F - z_i^{-2}\delta_{i\alpha}z_k\partial_{k\beta}F -$$

$$z_i^{-2}\delta_{i\beta}\delta_{k\alpha}\partial_k F + \left(z_i^{-1} - 1\right)\delta_{k\alpha}\partial_{k\beta}F - z_i^{-2}\delta_{i\beta}z_k\partial_{k\alpha}F +$$

$$\left(z_i^{-1} - 1\right)\delta_{k\beta}\partial_{k\alpha}F + \left(z_i^{-1} - 1\right)z_k\partial_{k\alpha\beta}F) \,,$$

$$\frac{\partial^3}{\partial t\partial z_\alpha\partial z_\beta}F\left(z,t\right)\mid_{z_i=1,i=1..M} = \tilde{A}_{\alpha\beta}\partial_\beta F + \sum_{k=1}^{M}\tilde{A}_{\alpha k}\partial_{k\beta}F + \tilde{A}_{\beta\alpha}\partial_\alpha F + \sum_{k=1}^{M}\tilde{A}_{\beta k}\partial_{k\alpha}F +$$

$$\sum_{k=1}^{M}2\,\tilde{\Gamma}_{\beta k}\partial_k F\delta_{\alpha\beta} - \tilde{\Gamma}_{\alpha\beta}\partial_\beta F - \sum_{k=1}^{M}\tilde{\Gamma}_{\alpha k}\partial_{k\beta}F - \tilde{\Gamma}_{\beta\alpha}\partial_\alpha F - \sum_{k=1}^{M}\tilde{\Gamma}_{\beta k}\partial_{k\alpha}F \,,$$

20

$$\boxed{\frac{d}{dt}F_{\alpha\beta} = \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) F_{k\beta} + \sum_{k=1}^{M} \left(\tilde{A}_{\beta k} - \tilde{\Gamma}_{\beta k} \right) F_{k\alpha} + \tilde{A}_{\alpha\beta}F_{\beta} + \tilde{A}_{\beta\alpha}F_{\alpha} \qquad (27)}$$
$$\boxed{-\tilde{\Gamma}_{\alpha\beta}F_{\beta} - \tilde{\Gamma}_{\beta\alpha}F_{\alpha} + 2\,\delta_{\alpha\beta} \sum_{k=1}^{M} \tilde{\Gamma}_{\beta k}F_{k}\,.}$$

This is the equation that we need. Later we will use it to reveal the action of the generators, that are hidden now in the coefficients $\tilde{A}_{ik}$. Before we deal with the generators, we will derive a general formula for the covariance matrix $\nu_{\alpha\beta}$ to see how different it is from the one above.

$$\nu_{\alpha\beta} \equiv\, <\tilde{q}_{\alpha}\tilde{q}_{\beta}> - <\tilde{q}_{\alpha}><\tilde{q}_{\beta}> = F_{\alpha\beta} - F_{\alpha}F_{\beta} + \delta_{\alpha\beta}F_{\alpha}\,, \qquad (28)$$

$$\frac{d}{dt}\nu_{\alpha\beta} = \frac{d}{dt}F_{\alpha\beta} - \left(\frac{d}{dt}F_{\alpha}\right)F_{\beta} - F_{\alpha}\frac{d}{dt}F_{\beta} + \delta_{\alpha\beta}\frac{d}{dt}F_{\alpha}\,. \qquad (29)$$

Now we insert the derivatives for F_{α} and $F_{\alpha\beta}$

$$\frac{d}{dt}\nu_{\alpha\beta} = \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) F_{k\beta} + \sum_{k=1}^{M} \left(\tilde{A}_{\beta k} - \tilde{\Gamma}_{\beta k} \right) F_{k\alpha} + \tilde{A}_{\alpha\beta}F_{\beta} + \tilde{A}_{\beta\alpha}F_{\alpha} \qquad (30)$$
$$-\tilde{\Gamma}_{\alpha\beta}F_{\beta} - \tilde{\Gamma}_{\beta\alpha}F_{\alpha} + 2\,\delta_{\alpha\beta} \sum_{k=1}^{M} \tilde{\Gamma}_{\beta k}F_{k} +$$
$$\sum_{k=1}^{M}(-\tilde{A}_{\alpha k}F_{k}F_{\beta} + \tilde{\Gamma}_{\alpha k}F_{k}F_{\beta} - \tilde{A}_{\beta k}F_{k}F_{\alpha} + \tilde{\Gamma}_{\beta k}F_{k}F_{\alpha}) +$$
$$\delta_{\alpha\beta} \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) F_{k}\,.$$

We want to get rid of the variables $F_{\alpha\beta}$ and write everything in terms of $\nu_{\alpha\beta}$ and $<q_{\alpha}>$. First we regroup the terms and then add and subtract the term

$$\sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) \delta_{k\beta}F_{\beta} \qquad (31)$$

to obtain

$$\frac{d}{dt}\nu_{\alpha\beta} = \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) \left(F_{k\beta} - F_k F_\beta + \delta_{k\beta} F_\beta \right) - \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) \delta_{k\beta} F_\beta + \tag{32}$$

$$+ \sum_{k=1}^{M} \left(\tilde{A}_{\beta k} - \tilde{\Gamma}_{\beta,k} \right) \left(F_{k\alpha} - F_k F_\alpha + \delta_{k\alpha} F_\alpha \right) - \sum_{k=1}^{M} \left(\tilde{A}_{\beta k} - \tilde{\Gamma}_{\beta k} \right) \delta_{k\alpha} F_\alpha +$$

$$+ \tilde{A}_{\alpha\beta} F_\beta + \tilde{A}_{\beta\alpha} F_\alpha - \tilde{\Gamma}_{\alpha\beta} F_\beta - \tilde{\Gamma}_{\beta\alpha} F_\alpha + 2\,\delta_{\alpha\beta} \sum_{k=1}^{M} \tilde{\Gamma}_{\beta k} F_k +$$

$$+ \delta_{\alpha\beta} \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) F_k$$

$$\frac{d}{dt}\nu_{\alpha\beta} = \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) \nu_{k\beta} + \sum_{k=1}^{M} \left(\tilde{A}_{\beta k} - \tilde{\Gamma}_{\beta k} \right) \nu_{k\alpha} - \tag{33}$$

$$- \tilde{A}_{\alpha\beta} F_\beta + \tilde{\Gamma}_{\alpha\beta} F_\beta - \tilde{A}_{\beta\alpha} F_\alpha + \tilde{\Gamma}_{\beta\alpha} F_\alpha +$$

$$+ \tilde{A}_{\alpha\beta} F_\beta + \tilde{A}_{\beta\alpha} F_\alpha - \tilde{\Gamma}_{\alpha\beta} F_\beta - \tilde{\Gamma}_{\beta\alpha} F_\alpha +$$

$$+ \delta_{\alpha\beta} \sum_{k=1}^{M} \tilde{\Gamma}_{\alpha k} F_k + \delta_{\alpha\beta} \sum_{k=1}^{M} \tilde{A}_{\alpha k} F_k$$

$$\boxed{\frac{d}{dt}\nu_{\alpha\beta} = \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} - \tilde{\Gamma}_{\alpha k} \right) \nu_{k\beta} + \sum_{k=1}^{M} \left(\tilde{A}_{\beta k} - \tilde{\Gamma}_{\beta k} \right) \nu_{k\alpha} + \delta_{\alpha\beta} \sum_{k=1}^{M} \left(\tilde{A}_{\alpha k} + \tilde{\Gamma}_{\alpha k} \right) < \tilde{q}_k > \qquad (34)}$$

A.2 The Generators

The generators constitute a submatrix of the matrix $\tilde{A}$:

$$\tilde{A} = \begin{pmatrix} 0_{\alpha\beta} & 0_{\alpha b} \\ \mathcal{G}_{a\beta} & A_{ab} \, . \end{pmatrix} \tag{35}$$

Here $0_{\alpha\beta}$ and $0_{\alpha b}$ are matrices with all elements zeros, where $\alpha, \beta = 1 \ldots n$, and $a, b = n+1 \ldots 3n$. The matrix $\mathcal{G}_{a\beta}$ contains the generators and thus is a matrix with time dependant elements. The matrix A_{ab} has constant elements which depend on the genetic network. From now on we make a distinction between Greek indices and Latin indices, so that we can rewrite the general time dependance equations in terms of generators. The Greek indices run along the DNA variables, whereas the Latin indices run through the mRNAs and proteins variables.

Lets specialize the equation (27) for Latin indices. We will split the summations using a generic Greek letter γ and a generic Latin letter g. We consider the number of DNA be constant in time and normalized to the value 1. As a consequence

$$F_{\gamma b} \equiv < \tilde{q}_\gamma \tilde{q}_b > -\delta_{\gamma b} < \tilde{q}_b >=< \tilde{q}_\gamma \tilde{q}_b >=< 1\tilde{q}_b >=< \tilde{q}_b >= F_b \tag{36}$$

In terms of Greek and Latin indices, the matrix $\tilde{\Gamma}$, looks like

$$\begin{pmatrix} 0_{\alpha\beta} & 0_{\alpha b} \\ 0_{a\beta} & \Gamma_{ab} \end{pmatrix} , \tag{37}$$

so

$$\frac{d}{dt}F_{ab} = \sum_\gamma (A_{a\gamma} - \Gamma_{a\gamma}) F_{\gamma b} + \sum_g (A_{ag} - \Gamma_{ag}) F_{gb} + \sum_\gamma (A_{b\gamma} - \Gamma_{b\gamma}) F_{\gamma a} + \sum_g (A_{bg} - \Gamma_{bg}) F_{ga} +$$

$$+A_{ab}F_b + A_{ba}F_a - \Gamma_{ab}F_b - \Gamma_{ba}F_a + 2\,\delta_{ab} \left(\sum_\gamma \Gamma_{b\gamma}F_\gamma + \sum_g \Gamma_{bg}F_g \right) ,$$

$$\frac{d}{dt}F_{ab} = (\sum_\gamma \mathcal{G}_{a\gamma}) < q_b > +(\sum_\gamma \mathcal{G}_{b\gamma}) < q_a > + \sum_g (A_{ag} - \Gamma_{ag}) F_{gb} + \sum_g (A_{bg} - \Gamma_{bg}) F_{ga} +$$

$$+A_{ab}F_b + A_{ba}F_a - \Gamma_{ab}F_b - \Gamma_{ba}F_a + 2\,\delta_{ab} \sum_g \Gamma_{bg}F_g .$$

We have to eliminate the sum of the generators. We use for this the equation for the mean, taking care that for DNA variables, $F_\gamma =< \tilde{q}_\gamma >= 1$

$$\frac{d}{dt}F_a = \sum_\gamma (A_{a\gamma} - \Gamma_{a\gamma}) F_\gamma + \sum_g (A_{ag}F_g - \Gamma_{ag}) F_g \tag{38}$$

$$= \sum_\gamma \mathcal{G}_{a\gamma} + \sum_g (A_{ag} - \Gamma_{ag}) F_g .$$

We obtain then:

$$\frac{d}{dt}F_{ab} = \left(\frac{d}{dt}F_a - \sum_g (A_{ag} - \Gamma_{ag}) F_g \right) F_b + \left(\frac{d}{dt}F_b - \sum_g (A_{bg} - \Gamma_{bg}) F_g \right) F_a +$$

$$+ \sum_g (A_{ag} - \Gamma_{ag}) F_{gb} + \sum_g (A_{bg} - \Gamma_{bg}) F_{ga} + A_{ab}F_b + A_{ba}F_a -$$

$$-\Gamma_{ab}F_b - \Gamma_{ba}F_a + 2\,\delta_{ab} \sum_g \Gamma_{bg}F_g ,$$

$$\frac{d}{dt}F_{ab} = \left(\frac{d}{dt}F_a\right)F_b + \left(\frac{d}{dt}F_b\right)F_a + \sum_g (A_{ag} - \Gamma_{ag})\,(F_{gb} - F_g F_b) + \sum_g (A_{bg} - \Gamma_{bg})\,(F_{ga} - F_g F_a) +$$

$$+ A_{ab}F_b + A_{ba}F_a - \Gamma_{ab}F_b - \Gamma_{ba}F_a + 2\,\delta_{ab}\sum_g \Gamma_{bg}F_g \ .$$

From the formula above, we see that a new variable appeared in a natural way:

$$X_{ab} = F_{ab} - F_a F_b \ . \tag{39}$$

The time evolution of this new variable is given by the equation:

$$\frac{d}{dt}X_{ab} = \sum_g (H_{ag}X_{gb} + H_{bg}X_{ga}) + A_{ab} < \tilde{q}_b > + A_{ba} < \tilde{q}_a > - \tag{40}$$

$$- \Gamma_{ab} < \tilde{q}_b > - \Gamma_{ba} < \tilde{q}_a > + 2\,\delta_{ab}\sum_g \Gamma_{bg} < \tilde{q}_g > ,$$

with

$$H_{ab} = A_{ab} - \Gamma_{ab} \ , \tag{41}$$

or

$$\frac{d}{dt}X_{ab} = \sum_g (H_{ag}X_{gb} + H_{bg}X_{ga}) + H_{ab} < \tilde{q}_b > + H_{ba} < \tilde{q}_a > + 2\,\delta_{ab}\sum_g \Gamma_{bg} < \tilde{q}_g > \ . \tag{42}$$

In what follows we will use a diagonal Γ matrix. For this case the equation simplifies to

$$\frac{d}{dt}X_{ab} = \sum_g (H_{ag}X_{gb} + H_{bg}X_{ga}) + A_{ab} < \tilde{q}_b > + A_{ba} < \tilde{q}_a > \ . \tag{43}$$

The meaning of the matrix X can be found if we write it in terms of the covariance matrix ν_{ab}.

$$X_{ab} = F_{ab} - F_a F_b = \nu_{ab} - \delta_{ab}F_a = \nu_{ab} - \delta_{ab} < \tilde{q}_a > \ . \tag{44}$$

Thus X measure the deviation of the stochastic process from a Poissonian process,

$$\nu_{ab} = \delta_{ab} < \tilde{q}_a > +X_{ab} . \tag{45}$$

Now it is easy to write everything in terms of the reduced state $q = (r_1, r_2, ..., r_n, p_1, p_2, ..., p_n)$. To do this we observe that $q_k = \tilde{q}_{k+n}$, $k = 1 ... 2n$. In other words, we subtract n from from each Latin index and keep the same notations for the variables. We use $i = a - n$, $j = b - n$, $k = g - n$.

First, for the equation for the mean we simplify a relation deduced before, (38)

$$\frac{d}{dt} F_i = \sum_\gamma (A_{i\gamma} - \Gamma_{i\gamma}) F_\gamma + \sum_k (A_{ik} F_k - \Gamma_{ik}) F_k \tag{46}$$

$$= \sum_\gamma \mathcal{G}_{i\gamma} + \sum_k (A_{ik} - \Gamma_{ik}) F_k .$$

Note that in the sum $\sum_\gamma \mathcal{G}_{i\gamma}$ only one term is nonzero for $i = 1 ... n$ and all terms are zero for $i = n = 1 ... 2n$. Indeed, each mRNA is controlled by only one generator:

$$\mathcal{G}_{i\gamma} = \delta_{i\gamma} g_i(t) . \tag{47}$$

The above formulas tell also that only the mRNA is under the control of the generator, not the proteins neither the DNA. To simplify the notation we will write

$$G_i(t) = g_i(t), \quad i = 1 ... n , \tag{48}$$

$$G_i(t) = 0, \quad i = n + 1 ... 2n . \tag{49}$$

The equation for the mean then simplifies to

$$\frac{d}{dt} < q_i >= \sum_k H_{ik} < q_k > +G_i(t) . \tag{50}$$

A.3 Solution to the Mean and Fluctuation equations

The two equations from the previous section can now be written using a matrix notation:

$$\frac{d}{dt} \mu = H\mu + G , \tag{51}$$

$$\frac{d}{dt} X = HX + XH^T + H \, diag(\mu) + diag(\mu) H^T + 2 diag(\Gamma \mu) , \tag{52}$$

where the column vector μ has the components $\mu_i = < q_i >$, and the matrix X is related with the covariance matrix ν by: $\nu_{ij} = \delta_{ij} < q_i > + X_{ij}$, $i, j = 1 \ldots 2n$. Here H^T is the transpose matrix of $H = A - \Gamma$ and $diag(\mu)$ has nonzero elements only on the principal diagonal: $diag(\mu)_{ij} = \delta_{ij}\mu_i$. We took care of the fact that X is a symmetric matrix, $X^T = X$.

The first equation in (51) has a column vector as an unknown, μ, and is easy to solve it if we use the Laplace transform

$$\mu(s) = \int_0^\infty e^{-st}\mu(t)dt .$$ (53)

The equation for the mean becomes

$$s\mu(s) - \mu_0 = H\mu(s) + G(s) ,$$ (54)

with μ_0 being the value of the mean number of molecules at time zero, when the generator was applied. Thus:

$$\mu(s) = (s - H)^{-1}(G(s) + \mu_0)$$ (55)

The next goal is to solve for X. The second equation in (51) is a matrix equation. To find a solution for X we transform the matrix equation into a vector equation. The transformation needed is ([3] page 244):

$$X \mapsto vec(X) ,$$ (56)

where the column vector $vec(X)$ contains the columns of the matrix X one on top of the next one, starting with the first column and ending with the last column. In index notations, the element X_{ij} of the matrix X gets into the line $i + m(j - 1)$ in $vec(X)$ if X is an $m \times m$ matrix.

The vec mapping has the useful property that

$$vec(HX) = (1 \otimes H)vec(X) ,$$ (57)

$$vec(XH) = (H^T \otimes 1)vec(X) ,$$ (58)

were 1 is the unit matrix and $A \otimes B$ is the tensor product of two matrices A and B. The matrix $A \otimes B$ is constructed by substituting each element a_{ij} of the matrix A by the matrix $a_{ij}B$.

The matrix equation for X becomes

$$\frac{d}{dt}vec(X) = (H \otimes 1 + 1 \otimes H)vec(X) + (H \otimes 1 + 1 \otimes H) vec(diag\,(\mu)) + 2vec(diag(\Gamma\mu)) \ . \quad (59)$$

The column vector $vec(diag\,(\mu))$ can be expressed in terms of the column vector μ:

$$vec(diag\,(\mu)) = L\mu \ , \quad (60)$$

were L is a lift matrix from a space of dimension of μ to the square of this dimension. The matrix L has the block structure

$$L = \begin{pmatrix} P_1 \\ \vdots \\ P_{2n} \end{pmatrix} \ , \quad (61)$$

where $2n$ is the number of rows in μ (n rows for mRNA and another n for proteins). The submatrices P_k, $k = 1...2n$ are $2n \times 2n$ square projection matrices, with all elements zero except one:

$$(P_k)_{ab} = \delta_{ak}\delta_{bk} \ . \quad (62)$$

As an example, for $n = 1$ we have 1 mRNA and 1 protein and the dimension of L is 4×2

$$L = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix} \ . \quad (63)$$

With the same lift matrix L we can write

$$vec(diag(\Gamma\mu)) = L\Gamma\mu \quad (64)$$

Denote now the Laplace transform of $vec(X)$ as V. We have, from 59,

$$sV(s) - V_0 = (H \otimes 1 + 1 \otimes H)V(s) + ((H \otimes 1 + 1 \otimes H)\,L + 2L\Gamma)\,\mu(s) \ , \quad (65)$$

27

$$V(s) = (s - 1 \otimes H - H \otimes 1)^{-1} ((1 \otimes H + H \otimes 1) L + 2 L\Gamma) (s - H)^{-1}(G(s) + \mu_0) \qquad (66)$$
$$+ (s - 1 \otimes H - H \otimes 1)^{-1} V_0 \; .$$

For a diagonal Γ

$$V(s) = (s - 1 \otimes H - H \otimes 1)^{-1} (A \otimes 1 + 1 \otimes A) L (s - H)^{-1}(G(s) + \mu_0) +$$
$$+ (s - 1 \otimes H - H \otimes 1)^{-1} V_0 \; ,$$

with μ_0 is the initial condition for the mean and V_0 the initial condition for $vec(X)$.

From the above formula (55) we see that the mean values are expressed in terms of the generators through the *mean transfer matrix*:

$$\frac{1}{s - H} \; . \qquad (67)$$

The interesting form, (66), is the *fluctuation transfer matrix*, that passes the time variation of the input generators into the time variation of $vec(X)$:

$$\frac{1}{s - 1 \otimes H - H \otimes 1} [(1 \otimes H + H \otimes 1) L + 2 L\Gamma] \frac{1}{s - H} \; .$$

For a diagonal Γ matrix this simplifies to

$$\frac{1}{s - 1 \otimes H - H \otimes 1} [(1 \otimes A + A \otimes 1) L] \frac{1}{s - H} \; . \qquad (68)$$

As an example, if H and Γ are 2 by 2 matrices,

$$H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \quad , \quad \Gamma = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \qquad (69)$$

we get:

$$s - 1 \otimes H - H \otimes 1 = \begin{bmatrix} -2\,h_{11} + s & -h_{12} & -h_{12} & 0 \\ -h_{21} & -h_{11} - h_{22} + s & 0 & -h_{12} \\ -h_{21} & 0 & -h_{11} - h_{22} + s & -h_{12} \\ 0 & -h_{21} & -h_{21} & -2\,h_{22} + s \end{bmatrix} \tag{70}$$

$$(1 \otimes H + H \otimes 1)\,L + 2\,L\Gamma = \begin{bmatrix} 2\,h_{11} + 2\,g_{11} & 2\,g_{12} \\ h_{21} & h_{12} \\ h_{21} & h_{1,2} \\ 2\,g_{21} & 2\,h_{22} + 2\,g_{22} \end{bmatrix} \tag{71}$$

$$((1 \otimes H + H \otimes 1)\,L + 2\,L\Gamma)\,\frac{1}{s - H} =$$

$$= \frac{1}{\Delta} \begin{bmatrix} 2\,h_{11}s - 2\,h_{11}h_{22} + 2\,g_{11}s - 2\,g_{11}h_{22} + 2\,g_{12}h_{21} & 2\,h_{12}h_{11} + 2\,h_{12}g_{11} + 2\,g_{12}s - 2\,g_{12}h_{11} \\ h_{21}\,(s - h_{22} + h_{12}) & h_{12}\,(h_{21} + s - h_{11}) \\ h_{21}\,(s - h_{22} + h_{12}) & h_{12}\,(h_{21} + s - h_{11}) \\ 2\,g_{21}s - 2\,g_{21}h_{22} + 2\,h_{21}h_{22} + 2\,h_{21}g_{22} & 2\,g_{21}h_{12} + 2\,h_{22}s - 2\,h_{11}h_{22} + 2\,g_{22}s - 2\,g_{22}h_{11} \end{bmatrix}$$

$$\Delta = s^2 - h_{22}s - h_{11}s + h_{11}h_{22} - h_{12}h_{21} \tag{72}$$

A.4 An autoregulatory gene with a periodically driven cofactor. Response of the system to an arbitrary input

One of the most fundamental regulatory motif in a genetic network is an autoregulatory gene through a negative feedback, [4]. We consider the case when the gene regulation is under the control of its own protein product and the protein activity is modulated by a cofactor. The equation for the mean is:

$$\frac{d}{dt} \begin{bmatrix} \langle r \rangle \\ \langle p \rangle \end{bmatrix} = \begin{bmatrix} -\gamma_r & -h \\ k_p & -\gamma_p \end{bmatrix} \begin{bmatrix} \langle r \rangle \\ \langle p \rangle \end{bmatrix} + \begin{bmatrix} k_0 + g(t) \\ 0 \end{bmatrix} \tag{73}$$

where the state is $q = (r, p)$. The cofactor, represented here by the term $g(t)$, is driven by the light generator. The cofactor modulates the mRNA mean number $\langle r \rangle$ through an additive coupling. For this case we use suggestive notations $X_{11} = X_{rr}$, $X_{12} = X_{rp}$ etc.. The Laplace transform of the quantities of interest, $(\langle r(t) \rangle, \langle p(t) \rangle, X_{rr}(t), X_{rp}(x), X_{pp}(t), g(t))$ are denoted by the same letter but the argument being the complex frequency s instead of the time t, like in

$$\langle r \rangle(s) = \int_0^\infty \langle r(t) \rangle e^{-st} dt . \tag{74}$$

The values of the mean number of molecules and their fluctuation, will depend on the internal parameters $\gamma_r, \gamma_p, h, k_p, k_0$ as well as on the external parameters of the generator $g(t)$. Two important natural parameters of the system play a significant role:

$$\omega_0^2 = hk_p + \gamma_r\gamma_p , \tag{75}$$

$$\omega_1 = \gamma_r + \gamma_p . \tag{76}$$

The mean number of molecules are connected to the generator through:

$$\begin{bmatrix} \langle r \rangle (s) \\ \langle p \rangle (s) \end{bmatrix} = \frac{1}{\Delta(s)} \begin{bmatrix} s + \gamma_p & -h \\ k_p & s + \gamma_r \end{bmatrix} \begin{bmatrix} g(s) \\ 0 \end{bmatrix} , \tag{77}$$

with

$$\Delta(s) = s^2 + s\omega_1 + \omega_0^2 . \tag{78}$$

The deviation from a Poisson process measured by the variable X is under the generator influence also:

$$\begin{bmatrix} X_{rr}(s) \\ X_{rp}(s) \\ X_{pr}(s) \\ X_{pp}(s) \end{bmatrix} = \frac{1}{\Delta_f(s)} \begin{bmatrix} -2h(s + 2\gamma_p)k_p(s + \gamma_p - h) & 2h^2(s + 2\gamma_p)(s + k_p + \gamma_r) \\ k_p(s + 2\gamma_p)(s + 2\gamma_r)(s + \gamma_p - h) & -h(s + 2\gamma_p)(s + 2\gamma_r)(s + k_p + \gamma_r) \\ k_p(s + 2\gamma_p)(s + 2\gamma_r)(s + \gamma_p - h) & -h(s + 2\gamma_p)(s + 2\gamma_r)(s + k_p + \gamma_r) \\ 2k_p^2(s + 2\gamma_r)(s + \gamma_p - h) & -2h(s + 2\gamma_r)k_p(s + k_p + \gamma_r) \end{bmatrix} \begin{bmatrix} g(s) \\ 0 \end{bmatrix} , \tag{79}$$

with

$$\Delta_f(s) = (s + \omega_1)\left(s^2 + s\omega_1 + \omega_0^2\right)\left(s^2 + 2s\omega_1 + 4\omega_0^2\right) \tag{80}$$

A.5 The step and the periodic stimuli

There are two cases of interest to us, a step stimulus and a periodic one.

For a step stimulus:

$$G(s) = \frac{G}{s} \ . \tag{81}$$

We consider that the system is in a steady state before we apply the step stimulus. The steady state is governed by the translation rate k_0. For a stable system ($\mathrm{Re}(\lambda_{1,2}) > 0$), the mean number of molecules decay exponentially to zero.

$$\langle r(t) \rangle = \frac{\gamma_p}{\lambda_1 \lambda_2} k_0 + \frac{\gamma_p}{\lambda_1 \lambda_2} G + \frac{(\lambda_1 - \gamma_p)}{\lambda_1 (\lambda_2 - \lambda_1)} G e^{-\lambda_1 t} + \frac{(\lambda_2 - \gamma_p)}{\lambda_2 (\lambda_1 - \lambda_2)} G e^{-\lambda_2 t} \tag{82}$$

$$\langle p(t) \rangle = \frac{k_p}{\lambda_1 \lambda_2} k_0 + \frac{k_p}{\lambda_2 \lambda_1} G + \frac{k_p}{\lambda_1 (\lambda_1 - \lambda_2)} G e^{-\lambda_1 t} + \frac{k_p}{\lambda_2 (\lambda_2 - \lambda_1)} G e^{-\lambda_2 t} \tag{83}$$

where λ_1 and λ_2 are the eigenvalues of H: $\Delta(s) = (s + \lambda_1)(s + \lambda_2)$,

$$\lambda_1 = 1/2\,\omega_1 - 1/2\,\sqrt{\omega_1{}^2 - 4\,\omega_0{}^2} \ , \tag{84}$$

$$\lambda_2 = 1/2\,\omega_1 + 1/2\,\sqrt{\omega_1{}^2 - 4\,\omega_0{}^2} \ . \tag{85}$$

For fluctuations we get also exponentially decaying responses to the step stimulus:

$$X_{rr}(t) = X_{rr,0} + X_{rr,\omega_1} e^{-\omega_1 t} + X_{rr,\lambda_1} e^{-\lambda_1 t} + X_{rr,\lambda_2} e^{-\lambda_2 t} + X_{rr,2\lambda_1} e^{-2\lambda_1 t} + X_{rr,2\lambda_2} e^{-2\lambda_2 t} \ , \tag{86}$$

$$X_{rp}(t) = X_{rp,0} + X_{rp,\omega_1} e^{-\omega_1 t} + X_{rp,\lambda_1} e^{-\lambda_1 t} + X_{rp,\lambda_2} e^{-\lambda_2 t} + X_{rp,2\lambda_1} e^{-2\lambda_1 t} + X_{rp,2\lambda_2} e^{-2\lambda_2 t} \ ,$$

$$X_{pp}(t) = X_{pp,0} + X_{pp,\omega_1} e^{-\omega_1 t} + X_{pp,\lambda_1} e^{-\lambda_1 t} + X_{pp,\lambda_2} e^{-\lambda_2 t} + X_{pp,2\lambda_1} e^{-2\lambda_1 t} + X_{pp,2\lambda_2} e^{-2\lambda_2 t} \ .$$

The coefficients from the above formulas are collected in the following matrices

$$\begin{bmatrix} X_{rr,0} \\ X_{rr,\omega_1} \\ X_{rr,\lambda_1} \\ X_{rr,\lambda_2} \\ X_{rr,2\lambda_1} \\ X_{rr,2\lambda_2} \end{bmatrix} = \begin{bmatrix} \frac{h k_p G (h - \gamma_p) \gamma_p}{\lambda_2{}^2 \lambda_1{}^2 \omega_1} + \frac{h k_p k_0 (h - \gamma_p) \gamma_p}{\lambda_2{}^2 \lambda_1{}^2 \omega_1} \\ 2 \frac{h k_p G (-2\gamma_p + \omega_1)(-\gamma_p + h + \omega_1)}{(-2\lambda_1 + \omega_1)(-\lambda_2 + \omega_1)(\omega_1 - 2\lambda_2)(-\lambda_1 + \omega_1)\omega_1} \\ -2 \frac{h k_p G (-\lambda_1 + 2\gamma_p)(\lambda_1 + h - \gamma_p)}{(\lambda_1 - 2\lambda_2)(-\lambda_2 + \lambda_1)\lambda_1{}^2 (-\lambda_1 + \omega_1)} \\ -2 \frac{h k_p G (2\gamma_p - \lambda_2)(\lambda_2 + h - \gamma_p)}{\lambda_2{}^2 (-\lambda_2 + \lambda_1)(-\lambda_2 + \omega_1)(2\lambda_1 - \lambda_2)} \\ \frac{h k_p G (-\lambda_1 + \gamma_p)(2\lambda_1 + h - \gamma_p)}{\lambda_1{}^2 (-\lambda_2 + \lambda_1)(2\lambda_1 - \lambda_2)(-2\lambda_1 + \omega_1)} \\ \frac{h k_p G (\gamma_p - \lambda_2)(2\lambda_2 + h - \gamma_p)}{\lambda_2{}^2 (-\lambda_2 + \lambda_1)(\lambda_1 - 2\lambda_2)(\omega_1 - 2\lambda_2)} \end{bmatrix} \ , \tag{87}$$

$$
\begin{bmatrix} X_{rp,0} \\ X_{rp,\omega_1} \\ X_{rp,\lambda_1} \\ X_{rp,\lambda_2} \\ X_{rp,2\lambda_1} \\ X_{rp,2\lambda_2} \end{bmatrix} = \begin{bmatrix} -\dfrac{k_p(k_0+G)\gamma_r\gamma_p(h-\gamma_p)}{\lambda_2{}^2\lambda_1{}^2\omega_1} \\[2ex] \dfrac{k_pG(\omega_1-2\gamma_r)(-2\gamma_p+\omega_1)(-\gamma_p+h+\omega_1)}{\omega_1(\omega_1-2\lambda_2)(-\lambda_2+\omega_1)(-\lambda_1+\omega_1)(-2\lambda_1+\omega_1)} \\[2ex] \dfrac{k_pG(-\lambda_1+2\gamma_r)(-\lambda_1+2\gamma_p)(\lambda_1+h-\gamma_p)}{\lambda_1{}^2(-\lambda_2+\lambda_1)(-\lambda_1+\omega_1)(\lambda_1-2\lambda_2)} \\[2ex] \dfrac{k_pG(-\lambda_2+2\gamma_r)(2\gamma_p-\lambda_2)(\lambda_2+h-\gamma_p)}{(-\lambda_2+\omega_1)(2\lambda_1-\lambda_2)(-\lambda_2+\lambda_1)\lambda_2{}^2} \\[2ex] -\dfrac{k_pG(-\lambda_1+\gamma_r)(-\lambda_1+\gamma_p)(2\lambda_1+h-\gamma_p)}{\lambda_1{}^2(-\lambda_2+\lambda_1)(2\lambda_1-\lambda_2)(-2\lambda_1+\omega_1)} \\[2ex] -\dfrac{k_pG(-\lambda_2+\gamma_r)(\gamma_p-\lambda_2)(2\lambda_2+h-\gamma_p)}{(-\lambda_2+\lambda_1)(\lambda_1-2\lambda_2)(\omega_1-2\lambda_2)\lambda_2{}^2} \end{bmatrix}, \tag{88}
$$

$$
\begin{bmatrix} X_{pp,0} \\ X_{pp,\omega_1} \\ X_{pp,\lambda_1} \\ X_{pp,\lambda_2} \\ X_{pp,2\lambda_1} \\ X_{pp,2\lambda_2} \end{bmatrix} = \begin{bmatrix} -\dfrac{k_p{}^2(k_0+G)\gamma_r(h-\gamma_p)}{\lambda_2{}^2\lambda_1{}^2\omega_1} \\[2ex] -2\,\dfrac{k_p{}^2G(-\gamma_p+h+\omega_1)(\omega_1-2\gamma_r)}{\omega_1(\omega_1-2\lambda_2)(-\lambda_2+\omega_1)(-\lambda_1+\omega_1)(-2\lambda_1+\omega_1)} \\[2ex] 2\,\dfrac{k_p{}^2G(\lambda_1+h-\gamma_p)(-\lambda_1+2\gamma_r)}{\lambda_1{}^2(-\lambda_2+\lambda_1)(\lambda_1-2\lambda_2)(-\lambda_1+\omega_1)} \\[2ex] 2\,\dfrac{k_p{}^2G(\lambda_2+h-\gamma_p)(-\lambda_2+2\gamma_r)}{(-\lambda_2+\lambda_1)(2\lambda_1-\lambda_2)(-\lambda_2+\omega_1)\lambda_2{}^2} \\[2ex] -\dfrac{k_p{}^2G(2\lambda_1+h-\gamma_p)(-\lambda_1+\gamma_r)}{\lambda_1{}^2(2\lambda_1-\lambda_2)(-\lambda_2+\omega_1)(-2\lambda_1+\omega_1)} \\[2ex] -\dfrac{k_p{}^2G(2\lambda_2+h-\gamma_p)(-\lambda_2+\gamma_r)}{\lambda_2{}^2(\lambda_1-2\lambda_2)(-\lambda_2+\lambda_1)(\omega_1-2\lambda_2)} \end{bmatrix}. \tag{89}
$$

For the periodic case with an input frequency ω and amplitude a, $g(t) = k_0 + a\cos(\omega t)$, (k_0 is a baseline not controlled by the exterior light input)

$$
g(s) = \frac{k_0}{s} + \frac{a}{s^2 + \omega^2}. \tag{90}
$$

We keep only the stationary solutions in the response (in practice we wait for the transients to become small enough)

$$
\langle r(t) \rangle = R_0 + R_1 e^{i\omega t} + R_1^* e^{-i\omega t}, \tag{91}
$$

$$
\langle p(t) \rangle = P_0 + P_1 e^{i\omega t} + P_1^* e^{-i\omega t}, \tag{92}
$$

$$
X_{rr}(t) = X_{r,0} + X_{r,1} e^{i\omega t} + X_{r,1}^* e^{-i\omega t}, \tag{93}
$$

$$
X_{rp}(t) = X_{rp,0} + X_{rp,1} e^{i\omega t} + X_{rp,1}^* e^{-i\omega t}, \tag{94}
$$

$$
X_{pr}(t) = X_{p,0} + X_{p,1} e^{i\omega t} + X_{p,1}^* e^{-i\omega t}. \tag{95}
$$

The star $*$ means complex conjugation. In terms of the parameters that constitutes the autoregulatory system we have:

$$R_0 = \frac{\gamma_p k_0}{\gamma_r \gamma_p + h k_p} \,, \tag{96}$$

$$P_0 = \frac{k_p k_0}{\gamma_r \gamma_p + h k_p} \,, \tag{97}$$

$$R_1 = 1/2 \frac{a\,(\gamma_p + i\omega)}{\omega_0{}^2 - \omega^2 + i\omega\,\omega_1} \,, \tag{98}$$

$$P_1 = 1/2 \frac{k_p\,a}{-\omega^2 + \omega_0{}^2 + i\omega\,\omega_1} \,, \tag{99}$$

$$X_{r,0} = \frac{k_0\,(h - \gamma_p)\,h k_p \gamma_p}{\omega_0{}^4 \omega_1} \,, \tag{100}$$

$$X_{rp,0} = \frac{k_0 \gamma_r \gamma_p k_p\,(\gamma_p - h)}{\omega_1 \omega_0{}^4} \,, \tag{101}$$

$$X_{p,0} = \frac{k_p{}^2 k_0\,(\gamma_p - h)\,\gamma_r}{\omega_0{}^4 \omega_1} \,, \tag{102}$$

$$X_{r,1} = \frac{-i a\,(-i\gamma_p + \omega + i h)\,(-\omega + 2\,i\gamma_p)\,h k_p}{(-\omega^2 + \omega_0{}^2 + i\omega\,\omega_1)\,(-\omega^2 + 2\,i\omega\,\omega_1 + 4\,\omega_0{}^2)\,(-\omega + i\omega_1)} \,, \tag{103}$$

$$X_{rp,1} = -1/2 \frac{a k_p\,(\omega - i\gamma_p + i h)\,(\omega - 2\,i\gamma_r)\,(\omega - 2\,i\gamma_p)}{(\omega^2 - \omega_0{}^2 - i\omega\,\omega_1)\,(\omega^2 - 4\,\omega_0{}^2 - 2\,i\omega\,\omega_1)\,(\omega - i\omega_1)} \,, \tag{104}$$

$$X_{p,1} = \frac{i a\,(-i\gamma_p + \omega + i h)\,(\omega - 2\,i\gamma_r)\,k_p{}^2}{(\omega^2 - \omega_0{}^2 - i\omega\,\omega_1)\,(\omega^2 - 2\,i\omega\,\omega_1 - 4\,\omega_0{}^2)\,(\omega - i\omega_1)} \,. \tag{105}$$

The impact of the natural (internal) frequencies ω_0 and ω_1 on the protein and mRN levels and fluctuation can be read out from the absolute values of the denominators of the mean and X:

$$\Delta = |det(i\omega - H)|^2 = \omega_0{}^2 \left((\omega^2 - \omega_0{}^2)^2 + \omega^2 \omega_1{}^2 \right) \,, \tag{106}$$

$$\Delta_f = |det(i\omega - 1 \otimes H - H \otimes 1)|^2 = 4\omega_1{}^2 \omega_0{}^2 \left(\omega_1{}^2 + \omega^2 \right)^2 \left((\omega^2 - 4\,\omega_0{}^2)^2 + 4\omega^2 \omega_1{}^2 \right) \,. \tag{107}$$

We observe that ω_0 is a resonance for the mean and X), whereas $2\omega_0$ is only for X.

Beside the ratios expressed by formulas (5) and (6) from the main paper, we can form different combinations between the periodic response variables that become useful for estimating the order of magnitude of the coefficients k, h, γ_r, γ_p (we consider the case when no experimental noise is present):

$$\frac{|R_1|^2}{|P_1|^2} = \frac{1}{k_p{}^2}\omega^2 + \frac{\gamma_p{}^2}{k_p^2} \,, \tag{108}$$

$$\frac{X_{r,0}}{X_{p,0}} = -\frac{h\gamma_p}{k_p \gamma_r} \,, \tag{109}$$

$$R_{1,\omega=0} = \frac{1}{2}\frac{\gamma_p}{\omega_0{}^2}\,a \,. \tag{110}$$

33

From the first relation we can estimate k_p and γ_p. From the second one we estimate the ratio $\frac{h}{\gamma_r}$. The third equation gives the variation of mRNA amplitude with the input amplitude a for small ω. From this relation we estimate ω_0. The mRNA degradation coefficient γ_r can now be obtained from

$$\gamma_r = \omega_0^2/(\gamma_p + \frac{h}{\gamma_r} k_p) \,. \tag{111}$$

Now we have h from h/γ_r. The last parameter k_0 comes from

$$R_0 = \frac{\gamma_p k_0}{\omega_0^2} \,. \tag{112}$$

There are other interesting ratios worth to be written down:

$$P_{1,\omega=0} = \frac{1}{2} \frac{k_p}{\omega_0^2} a \,, \tag{113}$$

$$\frac{|X_{p,1}|^2}{|X_{r,1}|^2} = \frac{h^2 \left(\omega^2 + 4\,\gamma_p^2\right)}{k_p^2 \left(\omega^2 + 4\,\gamma_r^2\right)} \,, \tag{114}$$

$$\frac{R_0}{P_0} = \frac{\gamma_p}{k_p} \,. \tag{115}$$

These relations can be used to further verify the validity of the model, once we estimated the parameters.

A.6 Fluctuation resonance

We want to find a driving frequency for which the fluctuations dominates over the mean values. For such a frequency the system will be in a pure fluctuation resonance. In such a situation the molecular noise can drive the cell out of its equilibrium state, which can have dramatic consequence on the cell fate. At the fluctuation resonance frequency, the deviation from a Poissonian process, measured by the quantity X, should be very high. To measure this deviation we consider the ratio of the fluctuation amplitude $|\,X_{p1}\,|$ over the mean amplitude $|\,P_1\,|$ (an analog of the Fano factor in frequency domain):

$$\frac{|\,X_{p1}\,|}{|\,P_1\,|} = \left(4\,k_p{}^2 \frac{\left(\omega^2 + (h - \gamma_p)^2\right)\left(\omega^2 + 4\,\gamma_r{}^2\right)}{\left((\omega^2 - 4\,\omega_0{}^2)^2 + 4\,\omega^2\omega_1{}^2\right)\left(\omega^2 + \omega_1{}^2\right)} \right)^{1/2} \,. \tag{116}$$

For systems for which $\omega_0 \gg \omega_1$ we can se a resonance for fluctuations but not for the mean values at the input frequency $\omega = 2\omega_0$. A plot of this ratio is presented in Fig.3. We notice that the

width and the hight of the resonance are inverse proportional. The parameters for which we see the resonance in Fig.3 doesn't belong to case we studied for step stimulus ($\lambda_{1,2}$ are real numbers there and complex here). The response to the step stimulus for systems that can enter into fluctuation resonance is a superposition of damped oscillations. Even in this situation the transients are gone after few periods.

A.7   The Genetic Network Spectral Function

The time response (mean and fluctuation) of the autoregulatory system to a step stimulus can be expressed in general as a sum of 6 terms

$$f_{exp}(t) = S_{exp,0} + S_{exp,1}e^{-\eta_1 t} + S_{exp,2}e^{-\eta_2 t} + S_{exp,3}e^{-\eta_3 t} + S_{exp,4}e^{-\eta_4 t} + S_{exp,5}e^{-\eta_5 t} . \tag{117}$$

Only three of these terms are present in the mean. For the purpose of the following analysis, we will consider only the case when all $\eta's$ are positive, which is equivalent with $\omega_1 > 2\omega_0$. The asymptotic response, of the same autoregulatory system, to a periodic stimulus has the form

$$f_{per}(t) = S_{per,0} + S_{per,1}e^{i\omega t} + S^*_{per,1}e^{-i\omega t} , \tag{118}$$

for both the mean and the fluctuation. The parameters of the system $k_p, h, \gamma_r, \gamma_p$ are hidden in the coefficients $S_{exp,i}$ or $S_{per,j}$, $i = 0, \ldots, 5$, $j = 0, 1$. For more complex genetic network, the time evolution of the measured quantity $f(t)$ can be expressed as

$$f(t) = \int_{x_1}^{x_2} S(x) K(xt) \, dx . \tag{119}$$

Here $S(x)$ is the *spectral function* that contains the information about the genetic network and $K(xt)$ is the kernel that depends only on the type of the stimulus (i.e. on the experimental design). Indeed, for an autoregulatory network, using the Dirac's δ-function, we have

$$f_{exp}(t) = \int_a^b S_{exp}(x) e^{-xt} dx , \tag{120}$$

$$S_{exp}(x) = \sum_{i=1}^6 S_{exp,i}\delta(x - \eta_i) , \tag{121}$$

$$K_{exp}(xt) = e^{-xt}. \tag{122}$$

The values a and b are chosen such that the spectrum $S_{exp}(x)$ is zero outside the interval $[a, b]$.

For the periodic stimulus, we have a similar representation for the spectral function $S(x)$ but the kernel is different

$$f_{per}(t) = \int_{-\Omega}^{\Omega} S_{per}(x) e^{ixt} dx , \tag{123}$$

$$S_{per}(x) = S_{per,0}\delta(x) + S_{per,1}\delta(x - \omega) + S^*{}_{per,1}\delta(x + \omega) , \tag{124}$$

$$K_{per}(xt) = e^{-ixt}. \tag{125}$$

with $\Omega > \omega$.

The topology of the genetic network is reflected in the spectral function $S(x)$. Given a set of measured data, first we have to recover the spectral function of the network and then from it the parameters of the network. If we lack a good model for the topology of the genetic network we cannot find the parameters of the network, but we can recover the spectral function $S(x)$ from the data (the kernel $K(xt)$ does not depend on the network). Thus different genetic networks can be compared using their spectral functions. However, the spectral function depends on the experimental design. We proved for the autoregulatory system that the spectral function S_{per} is much simpler than S_{exp}. We want to show that there is even a deeper difference between these two experimental designs. Namely, in the presence of experimental noise, it is much easier to recover S_{per} from the experimental data than S_{exp}. This phenomena appeared in other branches of science and in many different forms. To adapt it to biology, we noticed that a legitimate question from a molecular biologist is: instead of creating new assays to measure S_{per} why is not enough to increase the number of replicates to obtain an accurate S_{exp} ? We will prove that the number of replicates for S_{exp} growth exponentially with the accuracy. In what follows we collect and use for our specific problem, results form [5],[6].

In laboratory measurements, we don't have $f(t)$ for all values of t. Rather, we have samples of it at discrete time points. For the periodic stimulus, we measure $f(\tau n)$, where $n = 0, 1, \ldots, N$. As a working example, consider the samples of the mean of the mRNA, $r(n) \equiv \langle r(\tau n) \rangle, n = 0 \ldots N - 1, .$ The unknown spectrum $S_{per}(\omega)$ and $r(n)$ are related through the equation:

$$r(n) = \int_{-\Omega}^{\Omega} e^{in\tau\omega} S_{per}(\omega) d\omega . \tag{126}$$

There are three parameters in the problem: τ, Ω, N. The sampling parameter τ must be such

36

that the input frequency ω_{in} can be detected in the output data, that is $\tau \leq \pi/\omega_{in}$. The frequency Ω should be greater than the input frequency ω_{in}. There is no condition on the number of points N. Because we have a finite number N of measured data points, the spectrum $S_{per}(\omega)$ can only be approximated as a weighted sum of N functions $\Phi_k(\omega)$ (see Appendix 1 and [6])

$$\tilde{S}_{per}(\omega) = \sum_{k=0}^{N-1} s_k \Phi_k(\omega). \tag{127}$$

The functions $\Phi_k(\omega), k = 0 \cdots N - 1$ come from a eigenvalue problem for an $N \times N$ matrix (see Appendix 1 at the end of this Supporting Material). Now,the experimental noise will alter the coefficients s_k so the recovered spectrum will be:

$$\tilde{S}_{per}(\omega) = \sum_{k=0}^{N-1} \left(s_k + \frac{\epsilon_k}{\beta_k} \right) \Phi_k(\omega) , \tag{128}$$

where the ϵ_k are the noise coefficients. The numbers β_k, $k = 1 \cdots N - 1$, come from the same eigenvalue problem as before and they depend only on the parameters τ, Ω, N and not on the noise coefficients ϵ_k. Due to noise, we cannot use all N terms in (128), but only the first J_p, for which

$$\frac{1}{\beta_k} < \frac{s_k}{\epsilon_k}, \qquad k = 1, \cdots, J_p . \tag{129}$$

The right hand side of (129) is the Signal to Noise Ration (SNR) and for simplicity we will consider that is independent of the index k. The numbers β_k decrease as k increase and so the condition for the cutoff J_p is simple

$$\frac{1}{\beta_{J_p}} < SNR < \frac{1}{\beta_{(J_p+1)}} . \tag{130}$$

The exponential case can be developed parallel to the periodic case,[5]. The problem now reads like

$$r(n) = \int_a^b e^{-p_n\lambda} S_{exp}(\lambda) \, d\lambda . \tag{131}$$

Unlike for the periodic case, here a geometric sampling is optimum [5]

$$p_n = \frac{q}{a} \Delta^n, \ \ n = 1 \ldots N . \tag{132}$$

The limits a and b are chosen so that the spectrum is nonzero only inside $[a, b]$. For the periodic case we know the input frequency so we don't have to guess an interval $[a, b]$ as we have to do

for the exponential case. Only the ratio $\gamma = b/a$ is important, as we see by changing the variable $\lambda = ax$

$$r(n) = \int_1^\gamma e^{-ap_n x} S_{exp}(ax)\, a\, dx \ . \tag{133}$$

Similar to the periodic case, solving an eigenvalue problem we can find an N-dimensional approximation to the spectrum. Because of the experimental noise we can use only J_e degrees of freedom, not N :

$$\tilde{S}_{exp}(\lambda) = \sum_{k=1}^{J_e} \left(s_k + \frac{\epsilon_k}{\alpha_k} \right) \Psi_k(\lambda) \ . \tag{134}$$

Here the terms ϵ_k are due to random experimental errors. Again, the functions $\Psi_k(\lambda)$ and the numbers α_k come from an eigenvalue problem (different from the periodic one) and they don't depend on the noise but only on the parameters a, q, Δ, γ, N (actually, the numbers α_k do not depend on the parameter a, only $\Psi_k(\lambda)$ does.) The cutoff J_e is noise dependent and is given by

$$\frac{1}{\alpha_{J_e}} < SNR < \frac{1}{\alpha_{(J_e+1)}} \ . \tag{135}$$

The cutoffs J_p and J_e are of prime importance because they measure the number of degrees of freedom in the recovered spectrum. Desirable is that both cutoffs be as close as possible to the number of measurements, N, which is the case when the Signal to Noise Ratio (SNR) is high. Although the equations (130) and (135) look formally similar, they give completely different solutions to the cutoffs. This is a consequence of the different rate at which the numbers α_k and β_k decrease to zero which we will study in the next section.

A.8   The number of replicates

The SNR dictates how many spectral components are reliable and can be use to recover the spectrum. We can imagine that by using replicates we can improve the SNR and so the two cases will come close to each other. This is not true; actually we need an unrealistic number of replicates to keep even few components for the exponential case. Indeed, with the help of r replicates, the SNR increase to

$$SNR\sqrt{r} \ , \tag{136}$$

and the equations for the number of components J_e, J_p to enter into the recovered spectrum are

$$\frac{1}{\alpha_{J_e}} \le SNR\sqrt{r} < \frac{1}{\alpha_{J_e+1}} \ , \tag{137}$$

$$\frac{1}{\beta_{J_p}} \le SNR\sqrt{r} < \frac{1}{\beta_{J_p+1}} \ . \tag{138}$$

The plots , Fig 7, of the number of replicates r as a function the number of spectral components J_e or J_p reveal that using a periodic stimulus we can use many more spectral components to recover the spectrum. The number of replicate growth very fast in the exponential case (for $SNR = 10$ we need 269 replicates for 4 spectral components), whereas in the periodic case, the number of replicates stays low for many spectral components (only for the 17th component it raises to 14, with $SNR = 10$).

The source of such a discrepancy is that the eigenvalues α_k tend fast to zero as

$$\alpha_k{}^2 = \frac{\pi}{\cosh\left(\pi\,\xi_k\right)} \tag{139}$$

where ξ_k tends to infinity like a polynomial of degree at least one in k (there is no analytical formula for ξ_k). For the plotted example, $\gamma = 5, q = 1/20, \ \Delta = 60^{1/20}$

$$\alpha_0 \ = \ 7.66 \cdot 10^{-1} \ , \tag{140}$$

$$\alpha_1 \ = \ 3.28 \cdot 10^{-2} \ , \tag{141}$$

$$\alpha_2 \ = \ 1.02 \cdot 10^{-3} \ , \tag{142}$$

$$\alpha_3 \ = \ 1.74 \cdot 10^{-5} \ , \tag{143}$$

$$\alpha_{19} \ = \ 9.94 \cdot 10^{-29} \ . \tag{144}$$

For the periodic case the situation is much better. Here the numbers β_k depends only on the product $\tau\Omega$ and so is customary to introduce the parameter w through $2\pi w = \tau\Omega$. Then, for $w = 1/3$ for example, we get

$$\beta_1 \ = \ 0.99 \ , \tag{145}$$

$$\beta_2 \ = \ 0.99 \ , \tag{146}$$

$$\beta_3 \ = \ 0.99 \ , \tag{147}$$

$$\beta_4 \quad = \quad 0.99 \, , \tag{148}$$

$$\beta_{20} \quad = \quad 0.000084 \, , \tag{149}$$

There is no an general analytical formula for β_k but it was proven that the first $2Nw$ beta numbers are close to 1 with the rest of them decreasing fast to zero. The fact that the majority of the eigenvalues for the periodic case are 1 whereas the eigenvalues for the exponential case decrease fast to zero is the source of the difference between the two cases.
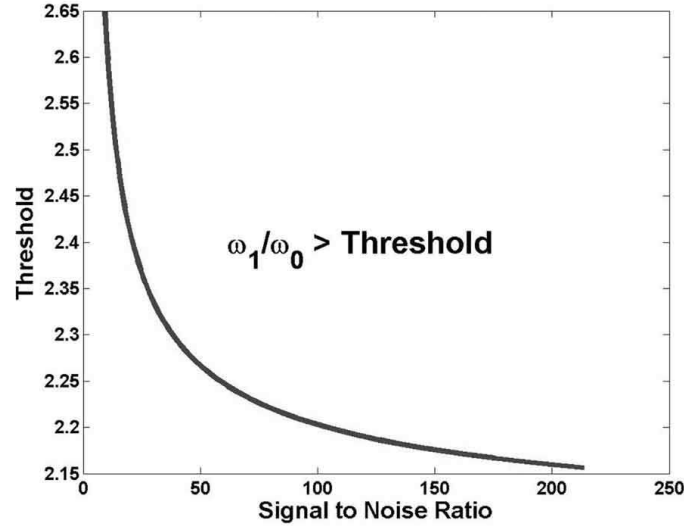

Figure 5: The Threshold as a function of SNR

Another interesting question is related to the resolution of the different exponentially decaying signals present in the output signal. For the periodic case we do not address this question, because the output signal has the same frequency as the input periodic signal (after the transients are gone). However, for the response to a step stimulus, the transients contain the information. To obtain this information we have to resolve the transient components.The resolution power depends on the Signal to noise Ratio (SNR). For example, to be able to resolve the decay rates λ_1 and λ_2 when they are real positive numbers, we need to have

$$\frac{\omega_1}{\omega_0} > Threshold(SNR) \, . \tag{150}$$

The Threshold as a function of SNR is plotted in the figure. Notice that we work with real $\lambda_{1,2}$ so $\omega_1 \geq 2\omega_0$ for all SNR.

40

Appendix 1. The eigenvalue problem for the periodic case

Recall that the measured quantities $r(n)$ for $n = 0 \ldots N - 1$, can be expressed as

$$r(n) = \int_{-\Omega}^{\Omega} e^{in\tau\omega} S_{per}(\omega)\, d\omega \ . \tag{151}$$

From the N data points we can find a N-dimensional approximation to the spectrum $S_{per}(\omega)$ solving the following singular value problem: find $V_k(\omega), v_k(n)$ and λ_k that satisfy

$$\mathcal{L}V_k = \lambda_k v_k \ , \tag{152}$$

$$\mathcal{L}^* v_k = \lambda_k V_k \ , \tag{153}$$

where the operator $\mathcal{L}$ and its conjugate $\mathcal{L}^*$ are

$$(\mathcal{L}f)(n) = \int_{-\Omega}^{\Omega} e^{in\tau\omega} f(\omega)\, d\omega \ , \tag{154}$$

$$(\mathcal{L}^* g)(\omega) = \sum_{n=0}^{N-1} e^{-in\tau\omega} g(n) \ . \tag{155}$$

The set $V_k(\omega)$ form an orthonormal basis in $L^2(-\Omega, \Omega)$ and $v_k(n)$ an orthonormal basis in the euclidian space E^N. In the V_k basis, the N-dimensional approximation to the spectrum reads like

$$S_{per}^N(\omega) = \sum_{k=0}^{N-1} \frac{r_k}{\lambda_k} V_k(\omega) \ , \tag{156}$$

where the coefficients r_k are obtained from the decomposition of the measured data $r(n)$

$$r(n) = \sum_{k=0}^{N-1} r_k v_k(n) \ . \tag{157}$$

The solution to the singular problem (152) can be reduced to the eigenvalue problem for the operator $\mathcal{L}\mathcal{L}^*$: find the eigenfunctions $v_k(n)$ and the eigenvectors λ_k^2 from

$$\sum_{m=0}^{N-1} \frac{\sin(\tau\Omega(m-n))}{\pi(m-n)} v_k(m) = \beta_k^2 v_k(n) \ , \tag{158}$$

where

$$\beta_k = \sqrt{\frac{\tau}{2\pi}} \lambda_k \ . \tag{159}$$

In this way, the solution to our problem is reduced to the diagonalization of an $N \times N$ matrix. This is the famous problem, [6]. We have two independent parameters τ and Ω. The eigenvalues of the problem (158) depends on w, defined as $2\pi w = \tau \Omega$. The first $2Nw$ eigenvalues are close to 1 with the rest of them close to zero. As a consequence, from (156) we see that we can keep only the first $2Nw$ terms, because the rest of them are highly amplified by the small values of the eigenvalues which is dramatic when the values r_k are corrupted by noise. We want than $2Nw$ to be close to N which case $w = 1/2$ and $\Omega = \pi/\tau$. This situation corresponds to a sampling parameter τ tuned for recovering the spectrum up to the frequency Ω. In general case when $\Omega \geq \pi/\tau$. The recovered spectrum, when noise is present will be than

$$\tilde{S}_{per}^{N}(\omega) = \sum_{k=0}^{N-1} \frac{r_k + \epsilon_k}{\lambda_k} V_k(\omega).$$
(160)

To connect with the notations from Section 10, denote $\Phi_k(\omega) = \sqrt{\tau/2\pi} V_k(\omega)$ and $s_k = r_k/\beta_k$.

Appendix 2. The eigenvalue problem for the step stimulus

The problem for the exponential decay responses was solved in [5]. The unknown spectrum S_{exp} and the measured data $r(n)$ are connected through the equation

$$r(n) = \int_{1}^{\gamma} e^{-a p_n x} S_{exp}(ax) \, a \, dx \,,$$
(161)

with $\gamma = b/a$ and $p_n = (q/a)\Delta^n, n = 1, \ldots, N$. Like for the periodic case, an N-dimensional approximation to the spectrum can be found from the solutions of two coupled equations:

$$\mathcal{K} U_k = \alpha_k u_k \,,$$
(162)

$$\mathcal{K}^* u_k = \alpha_k U_k \,,$$
(163)

where

$$(\mathcal{K}f)(n) = \int_{1}^{\gamma} e^{-a p_n x} f(x) \, dx \,,$$
(164)

$$(\mathcal{K}^* g)(x) = \sum_{n=1}^{N} w_n g(n) e^{-a p_n x} \,,$$
(165)

with the weights given by $w_n = p_n \ln(\Delta)$, see [5]. The unknowns are the functions $U_k(x)$ that form an orthonormal basis in $L^2(1, \gamma)$ and the functions $u_k(n)$ that form a basis in the euclidian space

$\mathbb{R}^N$ endowed with the scalar product

$$(g, h) = \sum_{n=1}^{N} w_n g(n) h(n) \ . \tag{166}$$

The N-dimensional approximation to the unknown spectrum can now be written as a decomposition in U_k basis as

$$S_{exp}^N(\lambda) = \frac{1}{a} \sum_{k=1}^{N} \frac{r_k}{\alpha_k} U_k(\lambda/a) \ , \tag{167}$$

with the components r_k obtained from decomposing the measured data r_n in the basis u_k

$$r_k = \sum_{n=1}^{N} w_n r(n) u_k(n) \ . \tag{168}$$

Similar to the periodic case, the eigenvalue problem to be solved now is

$$\sum_{m=1}^{N} \frac{\sqrt{w_n w_m}}{a} \frac{e^{-a(p_n+p_m)} - e^{-b(p_n+p_m)}}{p_n + p_m} \bar{u}_k(m) = \alpha_k^2 \bar{u}_k(n) \tag{169}$$

with $\bar{u}_k(n) = \sqrt{w_n}\, u_k(n)$. The matrix that is diagonalized in the problem (169) is a symmetrized version of $\mathcal{K}\mathcal{K}^*$ and so there is a scaling difference between u_k and $\bar{u}_k$. The eigenvalues α_k tend fast to zero as

$$\alpha_k{}^2 = \frac{\pi}{\cosh(\pi\,\xi_k)} \ , \tag{170}$$

where ξ_k tends to infinity like a polynomial of degree at least one. The recovered spectrum is

$$\tilde{S}_{exp}^N(\lambda) = \frac{1}{a} \sum_{k=1}^{N} \frac{r_k + \epsilon_k}{\alpha_k} U_k(\lambda/a) \ , \tag{171}$$

where the terms ϵ_k are due to random experimental errors.

In Section 10 we write the spectrum in terms of $\Psi_k(\lambda) = (1/a)U_k(\lambda/a)$ and $s_k = r_k/\alpha_k$.

Appendix 3. The eigenvalue problem for continuous measurements

We discussed the spectrum recovery from a finite number of data, which is the case of laboratory measurements. However, it is instructive to inspect the case when we know $f(t)$ from (119) for all t and in the limit for which $a = 0$, $b = \infty$ and $\Omega = \infty$. This problem was studied in [7]. As a bonus, we get an expression for the resolution of the exponential spectrum and a direct understanding of

the difference in the two eigenvalue problems presented in Appendix 3 and 4. The solution for the exponential case is in terms of the eigenvalues and eigenfunctions of the kernel $K(\mu t)$

$$\int_0^\infty K(\eta t) \Phi_n(\eta) \, d\eta = \Xi_n \Phi_n(t) \ . \tag{172}$$

The eigenfunctions form an orthogonal basis and both the measured data $f_{exp}(t)$ and the unknown $S_{exp}(\mu)$ can be decomposed as:

$$f_{exp}(t) = \sum_{k=1}^\infty f_{exp,k} \Phi_k(t) \ , \tag{173}$$

$$S_{exp}(\eta) = \sum_{k=1}^\infty S_{exp,k} \Phi_k(\eta) \ . \tag{174}$$

Now we have the relation between the spectrum and the measured data

$$S(\eta) = \sum_{k=1}^\infty \frac{f_{exp,k}}{\Xi_k} \Phi_k(\eta) \ , \tag{175}$$

with Ξ_k arranged in decreasing order $\Xi_1 > \Xi_2 > \ldots$. We see from this expression that if Ξ_k decrease to zero and the components of the measured data f_k are corrupted by noise, than the components with large k cannot be used to recover $S_{exp}(\eta)$. The function thus recovered $S_{exp}(\eta)$ has information just from the first components $f_{exp,k}$. Only if the eigenvalues don't decrease to zero we can use all the terms in the decomposition.

For the exponential decay problem (120) the eigenvalues form a continuous spectrum (k is a positive real number)

$$|\Xi_k|^2 = \frac{\pi}{\cosh(\pi k)} \tag{176}$$

For the periodic solution (123) with $\Omega = \infty$ (Fourier transform) the spectrum is discrete ($k = 0, 1, \ldots, \infty$)

$$\Xi_k = -i^k \sqrt{2\pi} \tag{177}$$

It is obvious the difference between the exponential decay situation (step stimulus) and the periodic response. In the former case the eigenvalues tend fast to zero whereas in the later case they never approach zero (they have a constant modulus one.)

We aim now to find the resolution limit for resolving the exponential decay problem [5]. For a given signal to noise ration SNR we want to find the minimum ratio of the exponential decay

rates η_i/η_{i+1} that can be resolved. Going a little deeper into the solution of the exponential decay response,[7] we find that the decomposition of the spectrum $g(\mu)$ is

$$S_{exp}(\eta) = \int_0^\infty a_k^+ \, \Xi_k^+ \, \Phi_k^+(\eta) \, dk + \int_0^\infty a_k^- \, \Xi_k^- \, \Phi_k^-(\eta) \, dk \tag{178}$$

where the eigenfunctions are

$$\Phi_k^+(\eta) = \frac{1}{\sqrt{k}\,\pi} \cos\left(k \ln(\eta) - \frac{\theta}{2}\right) \tag{179}$$

$$\Phi_k^-(\eta) = -\frac{1}{\sqrt{k}\,\pi} \sin\left(k \ln(\eta) - \frac{\theta}{2}\right) \tag{180}$$

with the angle θ expressed in terms of the Gamma function

$$\theta = angle\left(\Gamma\left(1/2 + ik\right)\right). \tag{181}$$

Due to noise, we can recover the components up to a maximum k_0, so we have all the components with $k < k_0$. For this reason, we only can resolve points on the axis η that are separated at a distance larger than the distance between two zeros of Φ_{k_0}. Due to the presence of $ln(\eta)$ in the argumet of the trigonometric function, the zeros are

$$\mu_m = e^{\frac{1/2\,\theta + m\pi}{k_0}} \tag{182}$$

To conclude, two decay rates η_a and η_b can be recovered from the measured data if

$$\frac{\eta_a}{\eta_b} > \frac{\eta_m}{\eta_{m+1}} = e^{\left(\frac{\pi}{k_0}\right)} \tag{183}$$

The value k_0 that is the index for the maximum eigenvalue recoverable from noise is given by comparing the signal to noise ratio with the eigenvalue

$$\frac{\cosh(\pi k_0)}{\pi} = SNR^2 \tag{184}$$

Applying (183) to the example we work with (84) we obtain the condition

$$\frac{\omega_1}{\omega_0} > 2 \cosh\left(\frac{\pi}{2\,k_0}\right) \tag{185}$$

References

[1] Thattai, M. & van Oudenaarden, A. (2001) *Proc. Natl Acad. Sci. USA* **98**, 8614-8619.

[2] van Kampen, N. G. (1992) *Stochastic Processes in Physics and Chemistry* (NorthHolland, Amsterdam).

[3] Horn, R.A. & Johnson, C.R. (1999) *Topics in Matris Analysis* (Cambridge University Press).

[4] Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I. et al. (2002) *Science* **298**, 799-804.

[5] Bertero, M., Brianzi, P. & Pike, E.R.(1985) *Proc.R.Soc.Lond.A* **398**, 23-44.

[6] Slepian, D. (1983) *SIAM Review* **25**, 379-393.

[7] McWhirter,J.G. & Pike, E.R. (1978) *J. Phys. A* **11**, 1729-1745 .